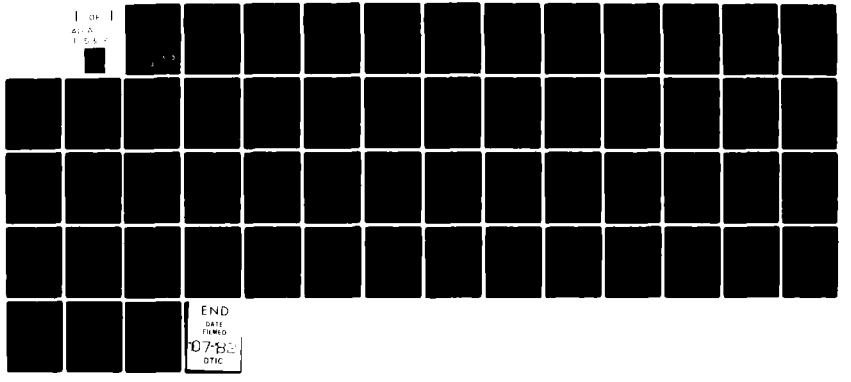
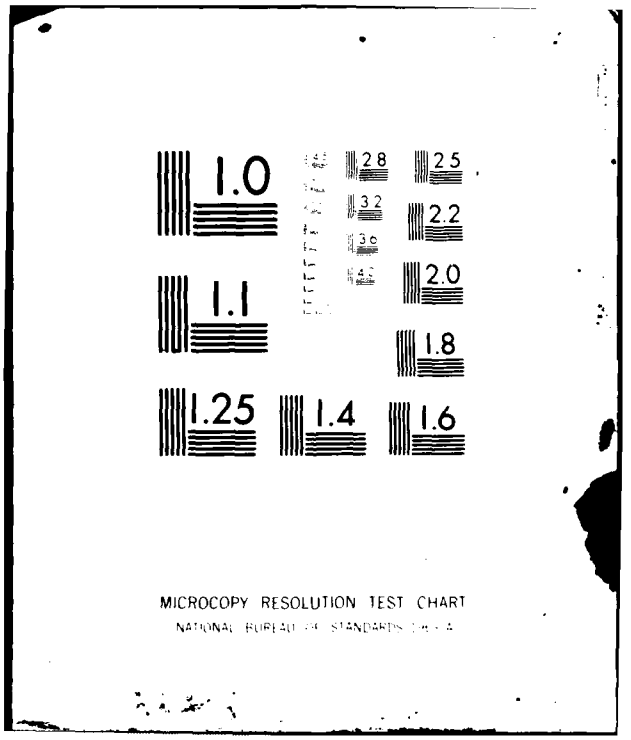
AD-A115 315 STANFORD UNIV CA DEPT OF STATISTICS F/G 12/1
ON THE PROBLEMS OF CONSTRUCTION AND STATISTICAL INFERENCE ASSOC--ETC(U)
FEB 82 A S GUPTA N00014-75-C-0442
UNCLASSIFIED TR-52 NL



END
DATE FILMED
07-82
DTIC


1.0
1.1
1.25
1.4
1.6
2.8
3.2
3.6
4.0
2.5
2.2
2.0
1.8
MICROCOPY RESOLUTION TEST CHART

(17)

ON THE PROBLEMS OF CONSTRUCTION AND STATISTICAL INFERENCE ASSOCIATED WITH A GENERALIZATION OF CANONICAL VARIABLES

BY

ASHIS SEN GUPTA

TECHNICAL REPORT NO. 52
FEBRUARY 1982

PREPARED UNDER CONTRACT N00014-75-C-0442
(NR-042-034)
OFFICE OF NAVAL RESEARCH

THEODORE W. ANDERSON, PROJECT DIRECTOR

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

ON THE PROBLEMS OF CONSTRUCTION AND STATISTICAL INFERENCE ASSOCIATED WITH A GENERALIZATION OF CANONICAL VARIABLES

by

ASHIS SEN GUPTA

TECHNICAL REPORT NO. 52

FEBRUARY 1982

PREPARED UNDER CONTRACT N00014-75-C-0442
(NR-042-034)
OFFICE OF NAVAL RESEARCH

Theodore W. Anderson, Project Director

Also issued as Technical Report No. 168 under National Science Foundation Grant MCS 78-07736 Department of Statistics, Stanford University.

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

0. INTRODUCTION

Canonical correlation analysis plays an important role in applied research since it provides a method of creating meaningfully bivariate random variables from two groups of random variables, each consisting possibly of a large number of components. The main purpose of this study is to investigate approaches to deal with the problems of construction and statistical inference which are posed when one tries to extend the theory of canonical correlation to deal with more than two sets of variables.

Anderson (1958) had suggested an interesting constructional approach for extending the theory of canonical variables. Though Steel (1951) and Kettenring (1971) have attempted to construct generalized canonical variables based on Anderson's approach, their results seem formidable for practical use--as have been speculated by Anderson. This paper presents a simple solution to Anderson's problem after modifying it with a useful constraint. Also problems of statistical inference related to generalized canonical variables are formulated and solutions to some of them are presented in this paper. As an application, important practical problems similar to the one posed by Gnanadesikan (1977), p. 77 can be formulated as a hypothesis testing problem and appropriate tests can be performed.

From a review of the existing literature on generalizations of canonical correlations, it was found that all the approaches to construct generalized canonical variables, except that due to Vinograde (1950) have the two natural properties - (1) they all reduce to the classical method

when the number of groups is only two and (2) the criterion for selection of generalized canonical variables optimizes some function of their correlation matrix. No work seems to have been done on the problems of statistical inference associated with the generalization of canonical variables.

Section 2 introduces a new approach for the construction of generalized canonical variables. The constraint of equi-correlation is discussed in the light of a variety of results and applications while the criterion of minimum generalized variance is shown to be of special interest through its statistical interpretation and examples of applications in a wide variety of fields. A brief formulation of the problems of statistical inference to be investigated in this research is also presented there.

In case of known population parameters, a derivation of the new generalized canonical variables is given in Section 3. Statistical estimation of these variables, when population parameters are unknown is given next and the interesting theory for the singular case concludes this section.

For the new generalized canonical variables, tests of hypotheses concerning their equi-correlation coefficient and generalized variance are of special interest. Using a large-sample approximation for the distribution of the new variables in case of normal population, various tests for the above hypotheses are proposed in Section 4. Some Properties of the above tests and their relationship with the Likelihood Ratio Test for correlation coefficient is also investigated in this section.

Tests for deciding whether the new generalized canonical variable of a particular size is as good as the one with a larger number of

elements are given in Section 5. Also, in this section tests for comparing several new generalized canonical variables in some special cases are given using isotonic regression techniques.

Finally, some discussions and topics for future research are presented in Section 7.

1. DEVELOPMENT OF GENERALIZED CANONICAL VARIABLES

1.1 Genesis:

A random vector of interest, in many situations, poses both analytical and economical problems if it has too many components. The objective creation of a new random vector with a much smaller number of components from the original one with a large number of components can often be of great help in applied research. The first principal component in effect seeks to determine a linear combination of the components of a p-dimensional random vector which explains the maximum variance among a set of all possible standardized linear functions. This procedure has been extended by Gnanadesikan and Wilk (1969) to search for a non-linear combination, giving rise to non-linear first principal component. So, a p-dimensional vector can be represented by just a scalar variable. Very often, the components of the original vector can be meaningfully grouped into several disjoint subvectors. In the case of two groups, Hotelling (1936) proposed canonical variables, where the first canonical variable seeks a two-component vector, each element being a standardized linear combination of the corresponding group, such that the correlation between any two such standardized linear combinations is maximum. The main advantage of principal component and canonical correlation analysis is that they reduce the problem of dimensionality and further, meaningfully try to explain the variation in the set to a great extent. The method of

canonical correlation has been widely applied in economics, marketing and in most of the behavorial and social sciences.

Hotelling noted the need for the extension of his canonical variables to the case of several groups. The object here is the same as above. However, the criterion for the choice of the variables need to be carefully defined.

For the case of k groups, it is desirable that the criterion for the choice of the variates, should be some function of their correlation matrix. Various such criteria and hence generalizations of canonical variables have been proposed.

The case of a single variable per set has been studied by many authors, though from different viewpoints, e.g. Horst (1936), and McKeon (1966).

Kettenring (1971) has considered the algebraic derivation of the generalized canonical variables for more than two sets with several variables per set. An expository discussion of generalizations of canonical variables is given in Sen Gupta (1981c).

2. INTRODUCTION TO NEW GENERALIZED CANONICAL VARIABLES

2.1 Definitions:

In the previous section it was seen that various attempts to generalize canonical variables have been made. Though, no formal definition of GCV seems to exist, in the spirit of the above attempts the following definition can be given.

Definition 1: Suppose a random vector $\underline{X}$ can be partitioned meaningfully into k disjoint groups or sub-vectors, $\underline{X}^{(1)},\ldots,\underline{X}^{(k)}$. The first GCV $\underline{Y}^{(1)}$ with $\underline{Y}^{(1)\prime}=[Y_1^{(1)},\ldots,Y_k^{(1)}]=[f_{11}(\underline{X}^{(1)}),\ldots,f_{1k}(\underline{X}^{(k)})]$, where f_{ij}s are real-valued functions, is a k-dimensional random variable, the components of which are chosen so as to optimize a criterion based on some function of their correlation matrix. The generalized higher order canonical variables $\underline{Y}^{(2)},\underline{Y}^{(3)},\ldots.$ are also k-dimensional random variables, the components of which are chosen so as to optimize the same criterion with some additional constraints imposed at each stage regarding the relationships among these variables.

Note, for all known methods f_{ij}s are linear functions. However, following the lines of non-linear principal component analysis, it is expected that non-linear GCVs can also be developed where the situation warrants the necessity of such a one.

In the present work, the search is for linear combinations

of the elements of each group, such that they are standardized and are equi-correlated. Formally, for the method to be presented here, we have

Definition 2: The first new GCV is the vector $\underline{Y}^{(1)'}=[Y_1^{(1)},\ldots,Y_k^{(1)}] = [\underline{\alpha}_1^{(1)'}\underline{X}^{(1)},\ldots,\underline{\alpha}_k^{(1)'}\underline{X}^{(k)}]$, where $\underline{\alpha}_i^{(1)}$, i=1,...,k are chosen such that the $Y_i^{(1)}$, i=1,...k are equi-correlated and the generalized variance of $\underline{Y}^{(1)}$ is minimum. The s-th order new GCV, $\underline{Y}^{(s)}$, s=2,3,... has components $Y_i^{(s)}$, i=1,...,k; $Y_i^{(s)} = \underline{\alpha}_i^{(s)'}\underline{X}_i$, where $\underline{\alpha}_i^{(s)}$s are chosen so that $Y_i^{(s)}$ are equi-correlated, $\underline{Y}^{(s)}$ is uncorrelated with $\underline{Y}^{(r)}$, r<s and generalized variance of $\underline{Y}^{(s)}$ is minimum.

2.2 The constraint of equi- correlation and the criterion of minimum generalized variance:

The motivations leading to seek such new GCVs are presented next. Note that, though $\underline{Y}^{(s)}$ has equi-correlated components, the original vector variable $\underline{X}$ may have any arbitrary dispersion matrix.

(A) The equi-correlation model - the constraint: The equi-correlation model not only has wide applications and interesting properties but as we will see, leads to simplifications for the analysis of new GCVs.

(A.a) Examples and uses: The equi-correlation model has been greatly used in various fields, e.g. by Votaw (1950) in medical experiments, Seal (1964) in the study of basic patterns of growth of grasshoppers, Gupta and Panchapakesan (1969) in ranking and selection procedures,etc. Various problems of statistical inference have been researched using this model, e.g. in the same year (1968) in Biometrika by Geisser and Desu, Glesser and also Han. Because of its numerous applications, extensive tables for equi-correlated multivariate normal c.d.f. were published by Milton (1963).

(A.b) Properties: The equi-correlation structure possesses some interesting properties. Two of them are discussed below.

(i) In many tests of Multivariate Analysis, e.g. tests in MANOVA, Profile Analysis, Growth Curve Analysis etc. if sample size from each population is not at least p, then in order that the mean square ratio in these Repeated Measurements Designs have exact F-distributions, the equi-correlation condition was thought to be necessary. However, Huynh and Feldt (1970) proved that this is a sufficient condition.

(ii) Another property follows from the following

Lemma: For any correlation matrix R, $R=(r_{ij})$, $|R| \leq [1+(k-1)r][1-r]^{k-1}$ with equality iff $r_{ij} = \sum_{i \neq j} r_{ij} / k(k-1) = r$, for all i,j and where R is a k x k matrix.

Proof: See Aitken, Nelson and Reinfurt (1968).

Note that using the constraint that the new GCVs are equi-correlated, we not only have great advantages from practical considerations, such as wide applications in a variety of fields and availability of tables for the probability integrals in case of multivariate normal population, but also from theoretical point of view, since now many tests in Multivariate Analysis can be performed on these new GCVs because they satisfy a sufficient condition for the validity of such test criteria. Further, we have the important advantage of working with a vector having a much smaller number of components as compared to the original one. Also, if the criterion is minimization of generalized variance, in view of property (ii) above, the method seems to utilize the minimax principle.

(B) The generalized variance - the criterion. The importance of generalized variance in Multivariate Analysis is well known. We review some interesting features about it below.

(B.a) The generalized variance as a measure of multi-dimensional scatter:

The generalized variance of a random vector variable is the determinant of the dispersion matrix of that variable. For the correlation matrix R, $|R|$ was termed 'Scatter coefficient' by Frisch in 1929. This definition of the generalized variance is a natural extension of the variance in the univariate case as a measure of dispersion.

Here we will be concerned with only one type of measures of dispersion, namely, the scatter of a random variable from a point of reference (and will thus disregard generalizations of other types like range etc.). Some of the following discussions are found in Mathai(1967).

If X' is a random variable, m is a fixed point of reference and X=X'-m, then a measure of dispersion for X' from m can be defined by the following metric D satisfying the following axioms:

K1. $D(X) \geq 0$, $D(X) = 0$ iff $X = 0$ almost surely.

K2. $D(aX) = |a| D(X)$, where a is a scalar quantity.

K3. $D(X+Y) \leq D(X)+D(Y)$ where Y'is another random variable and Y=Y'-m.

K4. $D(X) = 1$ if $|X| = 1$ almost surely.

A D(X) satisfying K1 to K4 can be called a measure of dispersion in X. From statistical considerations two more desirable properties of D(X)are

K5. $D(aX+b) = |a|^r D(X)$, where a,b and r ($r > 0$) are scalars.

K6. $D(X+Z) = D(X)+D(Z)$, where Z is another random variable independent of X.

If E denotes the operator mathematical expectation then an example of a measure of dispersion is given by $D(X)=[E|X|^r]^{1/r}$ for fixed $r \geq 1$. $E|X|^2 = \sigma^2$, satisfies K5 and K6, where $E(X') = m$.

If (X', Y') is a bivariate random variable, (m_1, m_2) is a point

of reference and $(X,Y) = (X'-m_1, X'-m_2)$ then the joint dispersion $D(X,Y)$ in (X,Y) can be defined by the following axioms:

L1. $D(X,Y)=D(Y,X)$.

L2. $D(aX,Y) = a\,D(X,Y)$ where a is a scalar quantity.

L3. $D(X+Z,Y) = D(X,Y)+D(Z,Y)$ where $Z=Z'-m_3$ and Z' is another random variable having a joint distribution with (X,Y).

L4. $D(X,X) = D_1(X)$ where $D_1(X)$ is an univariate measure of dispersion.

A desirable property here would be

L5. $D[A(X,Y)+B] = |A|^r D(X,Y)$, where $r(r>0)$ is a scalar, A,B are matrices with scalar elements and $A(X,Y)+B$ is a non-singular transformation of (X,Y).

The concept of covariance provides an example of a measure of joint dispersion.

Finally, consider a random vector $\underline{X}$ with k components. Let d_{ij} denote a measure of joint dispersion in the i^{th} and j^{th} variables. So the matrix $D_2 = (d_{ij})$ can be taken as a multivariate measure of joint dispersion in the k variables. Note that Σ provides an example of such a measure. A measure of generalized dispersion or multi-dimensional scatter would be a scalar quantity arising as a multivariate analogue of a univariate measure of dispersion. Hence such a measure can be defined as any norm of the matrix,i.e. should satisfy

M1. $||D|| > 0, ||D|| = 0$ iff D is a null matrix.

M2. $||aD|| = |a|\,||D||$ where a is a scalar quantity.

M3. $||C+D|| < ||C||+||D||$ where C is another matrix defined similarly.

M4. $||CD|| < ||C|| \cdot ||D||$

A desirable property here would be

M5. $||AD(\underline{X})+B|| = |A|^r ||D||$ where $r > 0$ is a scalar and A is non-singular.

Hence taking Σ, the usual variance-covariance matrix, for,D any norm of Σ can be taken as a measure of generalized dispersion.Largest characteristic root of Σ is an example of such a measure.

Note first that it is reasonable to expect that any scalar function of Σ, used as a measure of multi-dimensional statistical scatter should take into consideration the magnitude of the correlations among the variables. Secondly, van der Vaart (1965) has shown that the expected volume of the simplex formed by the k+1 random points in k dimensions or k random points and the mean vector is equal to the generalized variance, $|\Sigma|$. This is a natural generalization of the fact that the expected distance between two points or one point and the mean is the variance in one dimension. Thirdly, if the probability that a random point will lie in an ellipsoid of a k-dimensional distribution per unit volume is large, then the population is well concentrated about the mean. Finally, for a multivariate normal population, MLE of $|\Sigma|$ is the sample generalized variance as in the univariate case and distributional properties of the sample generalized variance has a marked similarity with that of the sample variance,e.g. for a random sample of size n, ns^2 is distributed as $\sigma^2\chi^2_{n-1}$ in the univariate case and $n^k s^{2k}$ as $\sigma^{2k}\chi^2_{n-1}\ldots\chi^2_{n-k}$ in the multivariate case with a vector variable of k components, where s^{2k} and σ^{2k} are the sample and population generalized variances respectively.

From the above properties and the works of Mathai (1967) entitled 'Dispersion theory' and Wilks (1967) entitled 'Multidimensional statistical scatter' the term generalized variance for the determinant of the dispersion matrix and its use as a measure of generalized dispersion seems to be very well coined (though it is a semi-norm).

(B.b) Applications of the generalized variance: Some situations are cited where the generalized variance can be used. Next, some examples where the use of the generalized variance has been advocated are presented.

Mathai (1967) pointed out that various statistical problems of estimation and tests of hypotheses can be looked upon as the study of properly defined measures of dispersion. Using the work of van der Vaart (1965), let $Q^2 = |B|^2/(k+1)!$, $B = \begin{bmatrix} 1 & \cdots & 1 \\ \underline{X}_1 & & \underline{X}_{k+1} \end{bmatrix}$ where $\underline{X}_i$ is $k \times 1$, $i = 1,\ldots,k+1$.

If $\underline{X} \sim F$, then $Q_X^2 \sim H(F)$, with mean as the generalized variance of $\underline{X}$. This is a very important and useful observation for testing purposes. Hence, any test for the equality of two location parameters can be used to provide a test for equal generalized variances. So, standard non-parametric tests for location can be used when the underlying distribution is unknown. Steyn (1978) has pointed out that testing the null hypothesis that the population mean vector $\underline{\mu}_r$ of the multivariate normal distribution with mean $\underline{\mu}_r$ and dispersion matrix Σ remains constant during the sampling process against the alternative that the mean vector varies during the process, is equivalent to testing the null hypothesis that the generalized variance $= |\Sigma|$ against the alternative hypothesis that the generalized variance $= |\Sigma^*|$ where $\Sigma^* = (I+2D/n)\Sigma$, $D = \sum_{r=1}^{n} \underline{\mu}_r \underline{\mu}_r' \Sigma^{-1}$, n is the sample size and $\underline{X}_r \sim N(\underline{\mu}_r, \Sigma)$, $r = 1,\ldots,n$.

Several practical examples are given below.

Example 1. (Press, 1972). A transistor produced by company j, C_j is characterised by a vector of k measurements, $\underline{X}_j : k \times 1$. Let $\underline{X}_j \sim N(\underline{\mu}, \Sigma_j)$, $j = 1,\ldots,r$. A purchaser would want to select the company with the property that he minimizes his chance of receiving a product that is sometimes of

unacceptably poor quality , that is choose that j for which $|\Sigma_j| = \min_{1\leq i\leq r}|\Sigma_i|$ Assuming a multivariate normal distribution,Gnanadesikan and Gupta (1970) and earlier Eaton (1967) have considered the problem of ranking the r underlying populations according to the magnitude of their generalized variances. Assuming that the loss function associated with the given ranking problem depends only on the r generalized variances $(\sigma_1^{2k},\ldots,\sigma_r^{2k})$, Eaton has shown that the decision rule which ranks the populations according to the sample generalized variances $(s_1^{2k},\ldots,s_r^{2k})$ possesses the following properties. It is (i) minimax within the class of all decision rules (ii) admissible within the class of decision rules which depend only on $(s_1^{2k},\ldots,s_r^{2k})$ and (iii) is the uniformly best decision rule among the class of rules which depend only on $(s_1^{2k},\ldots,s_r^{2k})$ and are invariant under permutations of $(s_1^{2k},\ldots,s_r^{2k})$.

Example 2. Though the definition of optimality becomes a problem in the multivariate situation, various authors have suggested the use of generalized variance in determining the optimum allocation of sampling units. Ghosh (1958) and Arvanitis and Afonja (1971) have proposed to determine the optimum sample size for the j-th strata, in stratified random sampling by minimizing the generalized variance of the sample means.

Example 3. Goodman (1966) has advocated the use of of the generalized variance to compare the overall variability of the populations of maize. It is apparent that a system of classification based on the overall similarities among the races is necessary both for the optimum usage of the race collections in breeding programs and for the study of the evolution of maize. Knowledge of the origin of the races against which values of the generalized variance can be compared, motivates here the

generalized variance as a reasonable measure of the overall variability and its use as a method of classifying the races of the maize.

(C) Minimization of the generalized variance - the choice: New GCVs are selected such that their generalized variance is minimum at each stage. The choice of minimization stems from two important considerations. Firstly, in the case of 2 groups, it is clear that maximizing the correlation of the 2 linear functions corresponding to the 2 groups, is equivalent to minimizing the generalized variance of thes two linear combinations. This property is extended to the case of k groups. Secondly, as has been illustrated in the various examples in sub-section (B) above, minimization of the generalized variance is widely used in many contexts and is of great practical use in various situations.

2.3 Formulation of the problems:

Broadly speaking, the two distinct problems here are those of construction and statistical inference related to GCVs. The present work deals with a random vector which can be meaningfully grouped into k disjoint sub-vectors. For the problem of construction, in its complete generality, one would face at least three problems of optimization - (i) How to select the groups in the absence of a given grouping (ii) How to select the components of the GCVs and (iii) How to decide on the optimal stage of stopping for higher order GCVs. (i) can possibly be solved partly by Cluster Analysis. For (ii) some available methods have been reviewed above. In the case of more than 2 groups, for (iii), as Kettenring has remarked for some of the methods above the situation is 'somewhat arbitrary'.

No work seems to be available for the problems of statistical inference related to the GCVs. Various problems would be interesting here. The distribution of GCVs needs to be derived. In case the form of the exact distribution is of not much use, suitable approximations and large sample distribution will be worth investigating. Several hypotheses develop here naturally from the defination of GCV and statistical tests are needed for these. Some such hypotheses are discussed next. (i) It would be natural to find out the extent of multidimensional statistical scatter explained by considering k groups. So, a test of hypothesis for a specific value of the generalized variance of the new GCV would be needed. Since one would want the new GCVs to have small scatter, the alternative hypothesis naturally should be of the less than type. So, we would test $H_0: |\Sigma_{\underline{Y}}| = \sigma_0^2$ against $H_1: |\Sigma_{\underline{Y}}| < \sigma_0^2$. (ii) Next, it would be interesting to see, if for the same dimension, regrouping of the variables in the original vector produces better result. For the new GCVs the related tests of hypotheses are - $H_{01}: |\Sigma_{\underline{Y}1}| = |\Sigma_{\underline{Y}2}|$ against $H_{11}: |\Sigma_{\underline{Y}1}| < |\Sigma_{\underline{Y}2}|$; $H_{02}: |\Sigma_{\underline{Y}j}|$s are all equal against $H_{12}: |\Sigma_{\underline{Y}j}|$s are in a given order. (iii) Also, it is reasonable to explore the possibility of whether the consideration of k groups leads to any substantial gain as compared to less number of groups which may be explicitly stated a-priori or which may be constructed by reasonably regrouping the variables in a fewer number of groups or amalgamating some of the original groups. The advantage here would be, that, in terms of the criterion, a GCV with a fewer number of components may perform as good as, if not better, than the original one with a larger number of components. This consideration may be, as in (ii), extended to the case of more than two GCVs with the

possibility of some of them to have same number of components. Since the criterion here must be comparable, the related tests of hypotheses become

H_{01}: $|\Sigma_{\underline{Y}1}|^{1/k_1} = |\Sigma_{\underline{Y}2}|^{1/k_2}$ against H_{11}: $|\Sigma_{\underline{Y}1}|^{1/k_1} < |\Sigma_{\underline{Y}2}|^{1/k_2}$ and

H_{02}: $|\Sigma_{\underline{Y}j}|^{1/k_j}$ are all equal against H_{12}: $|\Sigma_{\underline{Y}i}|^{1/k_i} < |\Sigma_{\underline{Y}j}|^{1/k_j}$, $k_i < k_j$

and a given ordering of generalized variances for new GCVs with equal number of components.

Note that in (ii) $\underline{Y}j$ referred to the new GCV obtained by the j-th mode of grouping the original random vector, j = 1,2,... while in (iii) $\underline{Y}j$ referred to the new GCV with number of components k_j, corresponding to the j-th mode of grouping, j = 1,2,... An example where such a situation may arise is provided by the following example [which corresponds to H_{01} in (ii)].

Example: (Thurstone and Thurstone, 1941). Horst, Kettenring and Seal (1964) discussed this example. Three (=k) sets of scores by several people on three batteries of three tests each (group size = 3 for each group) were obtained. The three tests in each battery were intended to measure, respectively, the verbal, the numerical and the spatial abilities of the persons tested. The correlation matrix of the original scores was given. Gnanadesikan (1977) commented "An interesting alternative analysis in this example would be to regroup the nine variables into three sets corresponding to the three abilities measured rather than the three batteries of tests." How profitable this alternative grouping would be in terms of multidimensional scatter can be judged by the test for H_{01} in (ii) above. Note that, since the $\underline{Y}j$ s would be correlated, such tests may not be very simple, even if the GCVs have a simple distribution.

No work seems to be available for the problems of statistical inference related to the GCVs. Various problems would be interesting here. The distribution of GCVs needs to be derived. In case the form of the exact distribution is of not much use, suitable approximations and large sample distribution will be worth investigating. Several hypotheses develop here naturally from the defination of GCV and statistical tests are needed for these. Some such hypotheses are discussed next. (i) It would be natural to find out the extent of multidimensional statistical scatter explained by considering k groups. So, a test of hypothesis for a specific value of the generalized variance of the new GCV would be needed. Since one would want the new GCVs to have small scatter, the alternative hypothesis naturally should be of the less than type. So, we would test $H_0: |\Sigma_{\underline{Y}}| = \sigma_0^2$ against $H_1: |\Sigma_{\underline{Y}}| < \sigma_0^2$. (ii) Next, it would be interesting to see, if for the same dimension, regrouping of the variables in the original vector produces better result. For the new GCVs the related tests of hypotheses are - $H_{01}: |\Sigma_{\underline{Y}1}| = |\Sigma_{\underline{Y}2}|$ against $H_{11}: |\Sigma_{\underline{Y}1}| < |\Sigma_{\underline{Y}2}|$; $H_{02}: |\Sigma_{\underline{Y}j}|$s are all equal against $H_{12}: |\Sigma_{\underline{Y}j}|$s are in a given order. (iii) Also, it is reasonable to explore the possibility of whether the consideration of k groups leads to any substantial gain as compared to less number of groups which may be explicitly stated a-priori or which may be constructed by reasonably regrouping the variables in a fewer number of groups or amalgamating some of the original groups. The advantage here would be, that, in terms of the criterion, a GCV with a fewer number of components may perform as good as, if not better, than the original one with a larger number of components. This consideration may be, as in (ii), extended to the case of more than two GCVs with the

possibility of some of them to have same number of components. Since the criterion here must be comparable, the related tests of hypotheses become

H_{01}: $|\Sigma_{\underline{Y}1}|^{1/k_1} = |\Sigma_{\underline{Y}2}|^{1/k_2}$ against H_{11}: $|\Sigma_{\underline{Y}1}|^{1/k_1} < |\Sigma_{\underline{Y}2}|^{1/k_2}$ and

H_{02}: $|\Sigma_{\underline{Y}j}|^{1/k_j}$ are all equal against H_{12}: $|\Sigma_{\underline{Y}i}|^{1/k_i} < |\Sigma_{\underline{Y}j}|^{1/k_j}$, $k_i < k_j$ and a given ordering of generalized variances for new GCVs with equal number of components.

Note that in (ii) $\underline{Y}j$ referred to the new GCV obtained by the j-th mode of grouping the original random vector, j = 1,2,... while in (iii) $\underline{Y}j$ referred to the new GCV with number of components k_j, corresponding to the j-th mode of grouping, j = 1,2,... An example where such a situation may arise is provided by the following example [which corresponds to H_{01} in (ii)].

Example: (*Thurstone and Thurstone*, 1941). Horst, Kettenring and Seal (1964) discussed this example. Three (=k) sets of scores by several people on three batteries of three tests each (group size = 3 for each group) were obtained. The three tests in each battery were intended to measure, respectively, the verbal, the numerical and the spatial abilities of the persons tested. The correlation matrix of the original scores was given. Gnanadesikan (1977) commented "An interesting alternative analysis in this example would be to regroup the nine variables into three sets corresponding to the three abilities measured rather than the three batteries of tests." How profitable this alternative grouping would be in terms of multidimensional scatter can be judged by the test for H_{01} in (ii) above. Note that, since the $\underline{Y}j$ s would be correlated, such tests may not be very simple, even if the GCVs have a simple distribution.

3. DERIVATION OF THE NEW GENERALIZED CANONICAL VARIABLES

3.1 Notations

Let $\underline{X}' = (\underline{X}_1', \ldots, \underline{X}_i', \ldots, \underline{X}_k')$, $\underline{X}_i' = (X_{i1}, \ldots, X_{ip_i})$, $i=1,\ldots,k$, $\sum_{i=1}^{k} p_i = p$, be a random vector variable meaningfully partitioned into k disjoint sub-vectors. Let $E(\underline{X})$ denote the expectation of $\underline{X}$ and since our interest lies only in the dispersion of $\underline{X}$, assume $E(\underline{X}) = \underline{0}$. Further, let $\Sigma(\underline{Z})$ and $|\Sigma(\underline{Z})|$ denote the dispersion matrix and the generalized variance of the random vector $\underline{Z}$, respectively. Then,

$$\Sigma(\underline{X}) = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} & \cdots & \Sigma_{1k} \\ \Sigma_{21} & \Sigma_{22} & \cdots & \Sigma_{2k} \\ \vdots & \vdots & \cdots & \vdots \\ \Sigma_{k1} & \Sigma_{k2} & \cdots & \Sigma_{kk} \end{bmatrix} \begin{matrix} p_1 \\ p_2 \\ \vdots \\ p_k \end{matrix} = \Sigma, \text{say}, \quad \Sigma(\underline{X}_i) = \Sigma_{ii}, \quad i = 1,\ldots k$$
$$\qquad\qquad p_1 \quad p_2 \quad \cdots \quad p_k$$

Denote the s^{th} new generalized canonical variable by $\underline{Y}^{(s)'} = (Y_1^{(s)}, \ldots, Y_k^{(s)})$, $Y_i^{(s)} = \underline{\alpha}_i^{(s)'} \underline{X}_i$, where $\underline{\alpha}_i^{(s)}$ is the coefficient vector of $\underline{X}_i$ at the s^{th} stage, $i = 1,\ldots,k$. Let, $\underline{\alpha}^{(s)'} = (\underline{\alpha}_1^{(s)'}, \ldots, \underline{\alpha}_k^{(s)'})$. Further, let $\underline{\alpha}' = (\underline{\alpha}'_1, \ldots, \underline{\alpha}'_k)$ denote the real but otherwise arbitrary coefficient vector which is used at the outset to derive the new generalized canonical variables $\underline{Y}$ at any stage and $\underline{\alpha}_i$s satisfy the condition that $\Sigma(\underline{Y})$ is equi-correlated, i.e.,

$$\Sigma(\underline{Y}) = \begin{bmatrix} \underline{\alpha}_1' \Sigma_{11}\underline{\alpha}_1 & \underline{\alpha}_1' \Sigma_{12}\underline{\alpha}_2 & \cdots & \underline{\alpha}_1' \Sigma_{1k}\underline{\alpha}_k \\ \underline{\alpha}_2' \Sigma_{21}\underline{\alpha}_1 & \underline{\alpha}_2' \Sigma_{22}\underline{\alpha}_2 & \cdots & \underline{\alpha}_2' \Sigma_{2k}\underline{\alpha}_k \\ \vdots & \vdots & & \\ \underline{\alpha}_k' \Sigma_{k1}\underline{\alpha}_1 & \underline{\alpha}_k' \Sigma_{k2}\underline{\alpha}_2 & \cdots & \underline{\alpha}_k' \Sigma_{kk}\underline{\alpha}_k \end{bmatrix} = \begin{bmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \cdots & \vdots \\ \rho & \rho & \cdots & 1 \end{bmatrix} = \Sigma_\rho \text{ say}$$

where ρ denotes the equi-correlation coefficient for Y_i's, i=1,...,k. Also, $\rho^{(s)}$ will denote the s^{th} new generalized canonical correlation coefficient.

3.2 Derivations of the new generalized canonical variables.

For the first new generalized canonical variable, we want to minimize $|\Sigma_\rho|$ subject to scale conditions, that, for Y_is, variances are all unity and they are equi-correlated. We use the method of Lagrange's undetermined multiplier. Let,

$$\phi = |\Sigma_\rho| + \sum_i \frac{\lambda_i}{k(k-1)} (\underline{\alpha}_i' \Sigma_{ii}\underline{\alpha}_i - 1) + \sum\sum_{i\neq j} \nu_{ij} \{ \sum\sum_{i\neq j} \underline{\alpha}_i' \Sigma_{ij}\underline{\alpha}_j - k(k-1)\underline{\alpha}_i' \Sigma_{ij}\underline{\alpha}_j \}$$

The last term introduces the condition of equi-correlation of the Y_is and is justified by the result that, $\bar{Z} = \sum_{i=1}^{m} \sum_{j=1}^{n} Z_{ij}/mn = Z_{st}$ for all s,t iff $Z_{ij} = Z_{kl}$ for all i,j,k,l. The λ_is and ν_{ij}s are Lagrange's undertermined multipliers, with $\nu_{ij} = \nu_{ji}$, $i \neq j$.

$$\frac{\partial \phi}{\partial \underline{\alpha}_t} = C(\rho) \frac{2}{k(k-1)} \sum_{j(j\neq t)} \Sigma_{tj}\underline{\alpha}_j + 2 \frac{\lambda_t}{k(k-1)} \Sigma_{tt}\underline{\alpha}_t$$

$$+ 2\left(\sum\sum_{i\neq j} \nu_{ij}\right) \sum_{j(j\neq t)} \Sigma_{tj}\underline{\alpha}_j - 2k(k-1) \sum_{j(j\neq t)} \nu_{tj}\Sigma_{tj}\underline{\alpha}_j = 0 \quad \ldots \quad \ldots(t.1)$$

for $t = 1,\ldots,k$ and write $t+1 = k+1$ as 1, and where $C(\rho) = \partial|\Sigma_\rho|/\partial_\rho$. We also used $\rho = \sum_{i\neq j} \underline{\alpha}_i'\Sigma_{ij}\underline{\alpha}_j/k(k-1)$ in the first term and the fact that $\underline{\alpha}_i'\Sigma_{ij}\underline{\alpha}_j = \underline{\alpha}_j'\Sigma_{ij}\underline{\alpha}_i$ for all $i,j = 1,\ldots,k$ in the last term. Multiplying (t.1) by $\underline{\alpha}_t'$, subtracting successive equations and using condition of equi-correlation, we get

$$\{k(k-1)\}^{-1}(\lambda_t - \lambda_{t+1}) + \{\sum_{j(j\neq t, j\neq t+1)} (\nu_{tj} - \nu_{t+1j}) + (\nu_{tt+1} - \nu_{t+1t})\}$$
$$\{-k(k-1)\rho\} = 0$$

or

$$\{k(k-1)\}^{-2}(\lambda_t - \lambda_{t+1}) = \rho\{\sum_{j(j\neq t, j\neq t+1)} (\nu_{tj} - \nu_{t+1j}) + (\nu_{tt+1} - \nu_{t+1t})\} \quad \ldots(t.2)$$

A solution to the set of equations defined by (t.2), $t = 1,\ldots,k$ is given by

$$\nu_{tj} = \nu_{t+1j} \text{ for all } j\neq t,\ t+1;\ \nu_{tt+1} = \nu_{t+1t} \text{ for all } t \text{ and}$$

$$\lambda_t = \lambda_{t+1} \text{ for all } t,\ t = 1,\ldots,k. \quad \ldots(3.2.1)$$

Using these values in (t.1) and λ as the common value of the λ_ts we get the simpler equation, [since (3.2.1) with $\nu_{ij} = \nu_{ji}$ gives ν_{ij}'s all equal]

$$C(\rho)\frac{2}{k(k-1)}\sum_{j(j\neq t)}\Sigma_{tj}\underline{\alpha}_j + 2\frac{\lambda}{k(k-1)}(\Sigma_{tt}\underline{\alpha}_t) = 0 \ldots \quad \ldots(t.3)$$

Multiplying the above equation by $\underline{\alpha}_t'$ we get, using the conditions of equi-correlation and variance of $Y_t = \underline{\alpha}_t'\Sigma_{tt}\underline{\alpha}_t = 1$, $\lambda = -(k-1)\rho C(\rho)$. Substituting this value of λ in (t.3) we finally get, assuming $C(\rho) \neq 0$.

$$\sum_{j(j\neq t)}\Sigma_{tj}\underline{\alpha}_j - (k-1)\rho\Sigma_{tt}\underline{\alpha}_t = 0 \ldots \quad \ldots(t.4)$$

In writing the abcve system of equations explicitly, we have

$$\begin{bmatrix} -(k-1)\rho\Sigma_{11} & \Sigma_{12} & \cdots & \Sigma_{1k} \\ \Sigma_{21} & -(k-1)\rho\Sigma_{22} & \cdots & \Sigma_{2k} \\ \vdots & \vdots & \cdots & \vdots \\ \Sigma_{k1} & \Sigma_{k2} & \cdots & -(k-1)\rho\Sigma_{kk} \end{bmatrix} \begin{bmatrix} \underline{\alpha}_1 \\ \underline{\alpha}_2 \\ \vdots \\ \underline{\alpha}_k \end{bmatrix} = 0 \qquad \ldots(1)$$

A non-trivial solution in ρ (which is necessary for a solution satisfying the condition that variances of Y_is are all unity) exists if and only if the coefficient matrix of the $\underline{\alpha}$ is singular. Writing this condition in a compact form we get,

$$|\Sigma_{od} - (k-1)\rho\Sigma_d| = 0 \qquad \text{or} \qquad |\Sigma - \lambda^*\Sigma_d| = 0 \qquad \ldots(2)$$

where

$$\Sigma_{od} = \begin{bmatrix} 0 & \Sigma_{12} & \cdots & \Sigma_{1k} \\ \Sigma_{21} & 0 & \cdots & \Sigma_{2k} \\ \vdots & \vdots & \cdots & \vdots \\ \Sigma_{k1} & \Sigma_{k2} & \cdots & 0 \end{bmatrix}, \quad \Sigma_d = \begin{bmatrix} \Sigma_{11} & 0 & \cdots & 0 \\ 0 & \Sigma_{22} & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & \Sigma_{kk} \end{bmatrix} \text{ and } \lambda^* = 1+(k-1)\rho$$

We may call Σ_d a diagonal super-matrix (a term due to Horst) and Σ_{od} an off-diagonal super-matrix with respect to Σ.

Note 1. The solution in (2) coincides with that of McKeon's and hence to that of various other authors as pointed out in Chapter I. The criterion and derivation were of course different here.

2. In order that the ρs obtained from (2) satisfy the condition of a correlation matrix, we must have that $-1/(k-1) \leq \rho \leq 1$. Here $\rho \leq 1$

by definition. Next, from Theorem 1 in Appendix we note that $\lambda^* \geq 0$, since Σ is positive semi-definite and assume that Σ_d is positive definite. [If Σ_d is positive semi-definite, then at least one of the Σ_{ii}'s must be singular. The assumption reasonably requires that at least within each group we do not have any redundant variables - those which bear linear dependence with other elements in the same group]. Hence, $\rho \geq \frac{-1}{k-1}$.

We solve the determinantal equations in (2) for the ρ-roots in terms of λ_i^*s and then by the one-one relationship, determine the corresponding ρ_is. Since our object is to minimize $|\Sigma_\rho|$ we choose that ρ_j for which $|\Sigma_{\rho_j}|$ is minimum. Now.

$\frac{\partial}{\partial\rho}|\Sigma_\rho| = -k(k-1)\rho(1-\rho)^{k-2} = C(\rho)$. So, $|\Sigma_\rho| \downarrow$ for $\rho<0$ & $\uparrow$ for $\rho>0$.

Hence, we need to consider both positive and negative ρ_is and then choose the one say $\rho^{(1)}$ for which $|\Sigma_\rho|$ is minimum. The coefficient vector $\underline{\alpha}^{(1)}$ corresponding to $\rho^{(1)}$ is then determined from (1) and the conditions of equicorrelations and unit variances for the components of $\underline{Y}^{(1)}$. Then, the first new GCV is given by $\underline{Y}' = \underline{Y}^{(1)}{}' = (\underline{\alpha}^{(1)}{}'\underline{X}_1, \ldots, \underline{\alpha}_k^{(1)}{}'\underline{X}_k)$ where $\underline{\alpha}^{(1)}{}' = (\underline{\alpha}_1^{(1)}{}', \ldots, \underline{\alpha}_k^{(1)}{}')$ is partitioned in a fashion corresponding to that of $\underline{X}$.

3. We had previously assumed that $C(\rho) \neq 0$, which means $\rho \neq 0$ and $\rho \neq 1$. From (2), note that if $|\Sigma_{od}| \neq 0$, then $\rho \neq 0$ and if $|\Sigma| \neq 0$ then $\rho \neq 1$ (and $\rho \neq -\frac{1}{k-1}$).

For higher order new generalized canonical variables, the constraints are next discussed and then the derivation is given below.

In case of canonical correlation analysis, i.e., k=2, $p_1 \leq p_2$ at the first stage one seeks the pair $(Y_1^{(1)}, Y_2^{(1)}) = \underline{Y}^{(1)}$, with the maximum possible correlation. The procedure may be pursued further until p_1 pairs of canonical variables have been determined. At the s^{th} stage, the s^{th} pair of canonical variables $\underline{Y}^{(s)}$ are found such that the corresponding s^{th} canonical correlation, $\rho^{(s)}$ is the maximum correlation attainable satisfying the conditions,

$$\text{corr}(Y_1^{(s)}, Y_1^{(t)}) = \text{corr}(Y_2^{(s)}, Y_2^{(t)}) = \text{corr}(Y_1^{(s)}, Y_2^{(t)})$$
$$= \text{corr}(Y_1^{(t)}, Y_2^{(s)}) = 0, \quad t = 1,\ldots,s-1.$$

Usually the first two conditions are sufficient, since they imply the latter ones and sometimes even only the first. Kettenring has considered in detail the above situations.

For GCVs, the same approach may be followed, introducing suitable restrictions such that the GCV for a particular stage are distinct from those of the previous stages. At the $(r+1)^{th}$ stage we invoke the restrictions,

$$\text{cor.}(Y_t^{(\ell)}, Y_j^{(r+1)}) = 0, \quad \ell = 1,\ldots,r; \; t,j = 1,\ldots,k$$

which is equivalent to the restrictions above for k=2, i.e., canonical variables. This type of restriction has been demonstrated by Kettenring to be equivalent to the restrictions

$$\text{corr.}(F_t^{(\ell)}, F_j^{(r+1)}) = 0, \quad \ell = 1,\ldots,r; \; t,j = 1,\ldots,k$$

where $F_t^{(r+1)}$ are the best fitting factors associated with the

$(r+1)^{th}$ stage fit of a k factor model like $\underline{Y}^{(1)} = \sum_{j=1}^{k} \underline{\xi}_j^{(1)} F_j^{(1)} + \underline{e}_1$ where the $\underline{\xi}_j^{(1)}$ are arbitrary non-null vectors, the $F_j^{(1)}$ are standardized random variables and $\underline{e}_1$ is a vector of error variables - using a criterion such as the one where the factors are determined so that $F_1^{(1)}$ is the most important, $F_2^{(1)}$ the second most important and so on. He has also pointed out interesting relationships of the above restrictions with the MAXVAR and MINVAR procedures in a factor analytic set-up.

At the $(r+1)^{th}$ stage using the above restriction, i.e.,

$\underline{\alpha}_i' \Sigma_{ij} \underline{\alpha}_j^{(\ell)} = 0 \quad \ell = 1,\ldots,r;\ i,j = 1,\ldots,k;$ or equivalently

$\underline{\alpha}_i' \Sigma_{ij} \underline{\alpha}_j^{(\ell)} - \underline{\alpha}_j' \Sigma_{ji} \underline{\alpha}_i^{(\ell)} = 0 = \underline{\alpha}_i' \Sigma_{ij} \underline{\alpha}_j^{(\ell)}$, we have,

$$\phi_{r+1} = \phi + \sum\sum\sum_{ijs} \theta_{ijs} \left(\underline{\alpha}_i' \Sigma_{ij} \underline{\alpha}_j^{(s)} - \underline{\alpha}_j' \Sigma_{ji} \underline{\alpha}_i^{(s)}\right) + \sum\sum_{is} \zeta_{is} \underline{\alpha}_i' \Sigma_{ii} \underline{\alpha}_i^{(s)}$$

where ϕ is the function defined in section 3.2. Hence, for t=1,...,k,

$$\frac{\partial \phi_{r+1}}{\partial \underline{\alpha}_t} = \frac{\partial \phi}{\partial \underline{\alpha}_t} + \sum\sum_{js} \theta_{tjs} \Sigma_{tj} \underline{\alpha}_j^{(s)} - \sum\sum_{is} \theta_{its} \Sigma_{ti} \underline{\alpha}_i^{(s)} + \sum_{s} \zeta_{ts} \Sigma_{tt} \underline{\alpha}_t^{(s)} = 0 \ldots \quad (t.5)$$

Pre-multiplying (t.5) by $\underline{\alpha}_t^{(\ell)}$, we get

$$\left(\sum_j \theta_{tj\ell} - \sum_i \theta_{it\ell}\right) \rho^{(\ell)} + \zeta_{t\ell} = 0 \ldots \qquad \ldots (t.6)$$

$$(t.6) - (t+1.6) \rightarrow \left[\sum_j (\theta_{tj\ell} - \theta_{t+1j\ell}) - \sum_i (\theta_{it\ell} - \theta_{it+1\ell})\right] \rho^{(\ell)} = \zeta_{t+1\ell} - \zeta_{t\ell}, \quad t=1,\ldots,k$$

A solution to these are obtained by taking θ_{tjs} all equal, ζ_{ts} all equal. But from (t.6) this means $\zeta_{ts} = 0$ for all t,s,t=1,...,k, s=1,...,r,; and using these values for θ_{tjs} and ζ_{ts} in (t.5) we get (t.1). Thus the previous arguments are applicable and we get (t.4) and (1). At the $(r+1)^{th}$ stage, we choose that root of (2) say $\rho^{(r+1)}$ which gives the $(r+1)^{th}$ lowest value of $|\Sigma_\rho|$ and call it the $(r+1)^{th}$ new generalized canonical correlation and the variable $\underline{Y}^{(r+1)'} = (\underline{\alpha}^{(r+1)'}\underline{X}_1,\ldots,\underline{\alpha}_k^{(r+1)'}\underline{X}_k)$ obtained from a corresponding coefficient vector as the $(r+1)^{th}$ new generalized canonical variable.

3.3 Sample new generalized canonical variables:

Since in practice both μ and Σ, the population mean and dispersion matrix are unknown, we need to estimate new GCVs from a sample of independent observations from the distribution of $\underline{X}$. To simplify matters we now need to assume that $\underline{X} \sim N_p(\mu,\Sigma)$. The procedure for non-normal populations may be pursued as done for canonical correlations, e.g. as in Rao (1973). Let $\underline{x}^u$, u = 1,...,n be a random sample of n observations from the population of $\underline{X}$ and let $\bar{\underline{x}}$ and S be the sample mean and dispersion matrix. Corresponding to the partition of $\underline{X}$, partition S. Then the MLE of Σ_{ij} and Σ are S_{ij} and S respectively. Now since $\{\underline{\mu},(\Sigma_d,\Sigma),\Lambda,A\}$, where $\Lambda = \text{diag}(\rho^{(1)},\ldots,\rho^{(p)})$, $A = (\underline{\alpha}^{(1)},\ldots,\underline{\alpha}^{(p)})$, is a single-valued function of $(\underline{\mu},\Sigma)$, it follows from Appendix A.2, that the MLE of $\rho^{(i)}$ is given by the corresponding characteristic root $r^{(i)}$, i = 1,...,p, of SS_d^{-1}. Also the MLE of $\underline{\alpha}^{(i)}$ is given by $\underline{a}^{(i)}$ satisfying $\underline{a}^{(i)'}\underline{a}^{(i)} = 1$, $(SS_d^{-1} - \underline{r}^{(i)}I)\underline{a}^{(i)} = \underline{0}$. Hence we get,

Theorem 1 The MLEs of the population non-zero new generalized canonical

correlations $\rho^{(1)},\ldots,\rho^{(q)}$, q being the number of non-zero roots, and the corresponding normalized coefficient vectors $\underline{\alpha}^{(1)},\ldots,\underline{\alpha}^{(q)}$, are given by respectively the ordered roots $r^{(1)},\ldots,r^{(q)}$ (i.e. $r^{(i)}$ gives the i^{th} minimum value of $|\Sigma_r|$ among the values of r, $r = r^{(j)}$, $j = 1,\ldots,p$) of the determinantal equation $|S - (k-1)rS_d| = 0$ and $\underline{a}^{(1)},\ldots,\underline{a}^{(q)}$, where $\underline{a}^{(i)}$s satisfy $(S-(k-1)r^{(i)}S_d)\,\underline{a}^{(i)} = \underline{0}$, $a_u^{(i)\prime}S_{uu}\underline{a}_u^{(i)} = 1$, $\underline{a}_u^{(i)\prime}S_{uv}\underline{a}_v^{(i)} = r^{(i)}$, $u \neq v$, $u,v = 1,\ldots,k$, $i = 1,\ldots,q$.

Note that, though unlikely, some of the non-zero $\rho^{(i)}$s may be known to be equal. The estimates have to be modified then. Standard derivations of MLE for the canonical correlations (e.g. Giri (1977),whose approach has been followed here) do not consider such cases. However, Anderson (1963) has given MLEs of characteristic roots in case of known multiple population roots for principal component analysis. The questions of under what situations some $\rho^{(i)}$s are equal and how to estimate them in those events seem to be interesting problems for further investigation. Also note that in the above derivation of MLEs we have used the Principle of Invariance of MLEs for non one-one functions advocated by Zehna(1966). This principle is stated in section A.2 of the Appendix.

3.4 The Singular Case:

Very often, when the vector variable under consideration has a large number of components, the dispersion matrix for the variable turns out to be singular. One is then faced with the problem of defining the generalized correlations and associated variables. In particular, for the generalized canonical correlations the following result overcomes that difficulty.

Theorem 4.4.1 The generalized canonical correlations, for the new method, are given by $\rho = (\rho^*-1)/(k-1)$ where ρ^*s are the non-zero roots of $|{}_k\Sigma_k\Sigma_d^- - \rho^* I| = 0$, ${}_k\Sigma_d^-$ being any g-inverse of ${}_k\Sigma_d$.

Proof: See proof of Theorem 1. Sen Gupta (1980).

Various other generalizations of canonical correlations are properly defined for the singular case by the application of the same Theorem 1 in Sen Gupta (1980).

4. STATISTICAL INFERENCE ASSOCIATED WITH A SINGLE SET OF NEW GENERALIZED CANONICAL VARIABLES

4.1 Distribution of the new generalized canonical variables:

The exact distribution of even the canonical variables with the special case of a multivariate normal population is complicated and is of doubtful practical importance (see e.g. Kshirsagar, 1972). The situation here is even more complex and so only the large sample distribution will be studied here. The coefficient vector $\underline{a}_i^{(s)}$, induced MLE of $\underline{\alpha}_i^{(s)}$, for $\underline{X}_i$ with $s = 1$ will be of particular importance, $i = 1,\ldots,k$. To simplify matters, we further assume that the population is multivariate normal because of the well known properties corresponding to linear functions of the components and the generalized variance of a random vector having a multivariate normal distribution. Since the estimated coefficient vector $\underline{a}_i^{(s)}$ is obtained by the principle of induced MLE, properties of such estimators need to be considered. The asymptotic efficiency of induced MLEs under certain regularity conditions is implicitly given in the proof by Zacks (1971) in Theorem 5.4.2 and there it is shown that $\hat{g}(\theta)$ can be taken as an induced MLE for $g(\theta)$ when $g(\theta)$ is a non one-one function. However, the case of the new GCVs is more complicated. We now have $\underline{\theta}$ as a $k(k-1)/2$ - component vector instead of a scalar and $\underline{g}(\underline{\theta})$ in addition to being non one-one can be scalar (for ρ) or a vector (for the coefficient

vector $\underline{\alpha}$). However the induced MLEs $\hat{\rho}^{(s)}$ and $\underline{a}^{(s)}$ of $\rho^{(s)}$ and $\underline{\alpha}^{(s)}$ respectively, are expected to be good approximations to corresponding population values under suitable comparisons. The validity of this approximation is shown below. We thus treat $\hat{\underline{\alpha}}^{(s)}$, in large samples, as the true coefficient vector itself and consider it as a vector of constants and for clarity write it as $\underline{\beta}^{(s)}$. $\underline{\beta}^{(1)}$ will be written as $\underline{\beta}$. Assuming that $\underline{X} \sim N_p(\underline{0},\Sigma)$, then $\underline{Y}' = (\underline{\beta}_1'\underline{X}_1,\ldots,\underline{\beta}_k'\underline{X}_k) = (Y_1,\ldots,Y_k) \sim N_k(\underline{0},\Sigma_{\rho^{(1)}})$, in large samples.

Assume that non-zero roots are all distinct. Then from a theorem due to Hurwitz (see Appendix) it follows that there exists a δ such that there will be exactly the same number of non-zero eigen values for A+E as for A when $||E|| < \delta$, where $||\ ||$ is a matrix-norm. Let λ' have the same rank among the eigen values of A+E as λ has among those of A. To make them unique, normalize the corresposding eigen vectors such that the first non-zero component of each is positive.

Consider without loss of generality the eigen-value $\lambda^{*(1)}$ of $\Sigma\Sigma_d^{-1}$ (corresponding to $\rho^{(1)}$) and the corresponding MLE, say g. By a standard theorem in complex analysis g is a continuous function of (S,S_d), because g is a continuous function of the coefficients obtained from the equation $|S - gS_d| = 0$ and in turn these coefficients are continuous functions of (S,S_d). The coefficient vector $\underline{a}^{(1)}$ corresponding to g, obtained by continuous operations on g and elements of S and S_d, is also a continuous function of (S,S_d). But $(S,S_d) \xrightarrow{P} (\Sigma,\Sigma_d)$ and so $\underline{a}^{(1)} \xrightarrow{P} \underline{\alpha}^{(1)}$, i.e. $\underline{a}^{(1)}$ is a consistent estimator of $\underline{\alpha}^{(1)}$ and so in large samples, the approximation considered above is meaningful.

4.2 Tests for equi-correlation coefficient $\rho^{(1)}$:

The value of $\rho^{(1)} \equiv \rho$ by itself, may be of interest. It may also be compared with the first canonical correlation when a sub-division of $\underline{X}$ into two groups also seems meaningful. Suppose $\underline{X}_1, \ldots, \underline{X}_n$ constitute an independent sample from $N_p(0,\Sigma)$ and let $\underline{Y}_1$ be the sample first new GCV. Then from section 4.1, in large samples $\underline{Y}_1 \sim N_k(\underline{0}, \Sigma_\rho)$. Let I and E be the identity matrix and the matrix with all elements equal to unity. Then,

$\Sigma_\rho = (1-\rho)I+\rho E$; $\Sigma_\rho^{-1} = (1-\rho)^{-1}I-\rho(1-\rho)^{-1}\{1+(k-1)\rho\}^{-1}E = (c_{ij})$, where

$c_{ii} = \{1+(k-2)\rho\}/(1-\rho)\{1+(k-1)\rho\}$ and $c_{ij} = -\rho/(1-\rho)\{1+(k-1)\rho\}$, $i \neq j$.

Hence the density function for non-singular Σ_ρ can be written as,

$$f(\underline{Y}_1;\rho) = \frac{1}{(2\pi)^{k/2}|\Sigma_\rho|^{1/2}} \exp\left[-\frac{1}{2}\left\{\frac{(\Sigma y_i^2)}{(1-\rho)} + \frac{(\Sigma y_i)^2(-\rho)}{(1+(k-1)\rho)(1-\rho)}\right\}\right] \ldots (4.2.1)$$

$$-\infty < y_i < \infty,\ i = 1,\ldots,k \text{ and } \rho \neq 1,\ \neq -1/(k-1).$$

The above representation is particularly useful because it shows that,

(i) there does not exist any one-dimensional sufficient statistic for ρ

(ii) $\Sigma (y_i - \bar{y})^2$ and $\bar{y} = \Sigma y_i/k$ are independent and

(iii) the part of the exponent within the second bracket is monotonically decreasing in ρ with positive probability.

We want to test H_0: $\rho = \rho_0$ against H_1: $\rho <(>) \rho_0$ or against H_2: $\rho \neq \rho_0$.

Likelihood ratio test: For testing H_0 against H_2 the LRT is derived below. We also write $\rho = \rho^{(1)}$ for simplicity, as stated in the outset and also suppose the original sample of size n is split into m independent subsamples, giving rise to m independent first new GCVs. The LRT can be performed even if m = 1. The likelihood function can be written from (1) above easily and differentiating this with respect to ρ and equating the derivative to zero we have,

$$g(\rho) = (k-1)k\rho(1-\rho)\{1+(k-1)\rho\}+\Sigma\Sigma_{ji} y_{ij}^2\{1+(k-1)\rho\}^2+\Sigma_j(\Sigma_i y_{ij})^2\{1+(k-1)\rho^2\} = 0,$$

where $\underline{y}_j' = (y_{1j},\ldots,y_{kj})$ is the sample first new GCV obtained from the j^{th} sub-sample, $j = 1,\ldots,m$. This is a cubic equation in ρ, two of whose roots may be complex. Now, $g\{-1/(k-1)\}$, $g(1)$ and $g(0)$ are all positive (with probability one). Thus, it is not very obvious from $g(\rho)$ that an admissible solution of ρ, say $\hat{\rho}$, where $-1/(k-1)<\hat{\rho}<1$, will always exist. [However, in various cases admissible solutions may exist, e.g. if $\Sigma\Sigma y_{ij}^2<1/2$, and $\Sigma(\Sigma y_{ij})^2<1/2$ then $g(-1/k)$ is negative. Hence, an admissible solution here lies in $(-1/(k-1),-1/k]$.] However, considering the likelihood function directly it can be shown that there exist at least one such admissible solution. Further, in cases of several admissible solutions (at most three), by principle of Maximum Likelihood, we choose as the MLE of ρ, that which corresponds to the largest value of the Likelihood function and call it $\hat{\rho}$. Thus we get the following

Theorem 1: Let $\underline{Y} \sim N_k(\underline{0},\Sigma_\rho)$. Then if $\underline{Y}_1,\ldots,\underline{Y}_m$ constitute an independent random sample from the above population, the Likelihood Ratio Test for testing H_0: $\rho=\rho_0$ against the alternative H_2: $\rho\neq\rho_0$, is given by

$$\text{Reject } H_0 \text{ iff } \lambda = \{|\Sigma_{\hat{\rho}}|/|\Sigma_{\rho_0}|\}^{m/2} \exp[-\tfrac{1}{2}\{a\{f_1(\rho_0)-f_1(\hat{\rho})\}+b\{f_2(\rho_0)-f_2(\hat{\rho})\}\}]$$

$$< K, \text{ a constant}$$

where $\hat{\rho}$ is the MLE of ρ, $a = \Sigma\Sigma_{ji} y_{ij}^2$, $b = \Sigma_j(\Sigma_i y_{ij})^2$, $f_1(\rho) = 1/(1-\rho)$, $f_2(\rho) = -\rho/\{1+(k-1)\rho\}(1-\rho)$ and K is a constant to be determined so that the level of the test meets the specified value.

Under H_o, for large m, $-2\ln\lambda \sim \chi_1^2$ d.f.

The exact distribution of $\hat{\rho}$ and the LRT seem quite complicated.

Due to the above difficulties in LRT a new test based on the Best Unbiased Estimator of ρ has been considered in Sen Gupta (1981b). The exact null and non-null distribution of the test statistic has been obtained in terms of Kummer's function and the test has been shown to

unbiased against one-sided alternatives. The distribution in (5.2.1) is also interesting because it constitutes a practical example of a member of Efron's curved exponential family. For detailed discussion the reader is referred to Sen Gupta (1981b).

4.3 Tests for a specified value of the generalized variance of the new generalized canonical variables.

Since the criterion for optimization is the generalized variance, a test for generalized variance is needed here to judge the performance of the new GCVs. The Likelihood Ratio Test for generalized variance, for the special structure of the dispersion matrix here, admists of some simplifications. If, $|\Sigma_\rho| = \sigma_0^2 < 1$, then there are precisely two real distinct solutions, $\rho_2 > 0 > \rho_1$, say.

Then (using MLE of ρ with $m = 1$ in Section 4.2) we have, by an application of the results in Sen Gupta (1981b),

Theorem 1. The Likelihood Ratio Test for $H_0: |\Sigma_\rho| = \sigma_0^2$ against $H_1: |\Sigma_\rho| \neq \sigma_0^2$ is given by,

$$\text{Reject } H_0 \text{ iff } \lambda = (|\Sigma_{\hat{\rho}}|/\sigma_0)^{k/2} \exp\{-\tfrac{1}{2}\, \underline{Y}'(\Sigma_{\rho^*}^{-1} - \Sigma_{\hat{\rho}}^{-1})\underline{Y}\} < C$$

where $\hat{\rho}$ is MLE of ρ, ρ^* is such that, $f(\underline{Y}_1, \rho^*) = \max_{\rho_1, \rho_2} f(\underline{Y}_1; \rho)$ and C is a constant to be determined such that the test has the desired level.

If the original sample is split into m independent subsamples, so that $\underline{Y}_j$, $j = 1, \ldots, m$ are new GCVs based on the j^{th} subsample, then,

$$\text{Reject } H_0 \text{ iff } \lambda_m = (|\Sigma_{\hat{\rho}}|/\sigma_0)^{mk/2} \exp\{-\tfrac{1}{2} \sum_{j=1}^{m} \underline{Y}_j'(\Sigma_{\rho^{**}}^{-1} - \Sigma_{\hat{\rho}}^{-1})\underline{Y}_j\} < C'$$

where ρ^{**} is such that $\prod_1^m f(\underline{Y}_j, \rho^{**}) = \max_{\rho_1, \rho_2} \prod_1^m f(\underline{Y}_j, \rho)$, $\hat{\rho}$ is the MLE of ρ and under H_0, for large m, $-2 \ln \lambda_m \sim \chi_1^2$.

For other interesting tests for the generalized variance, $|\Sigma_\rho|$ see Sen Gupta (1981b).

5. COMPARISON OF SEVERAL NEW GENERALIZED CANONICAL VARIABLES

5.1 Tests for equality of multidimensional scatter of two new GCVs:

As in Section 4 we will consider only the first new GCV. The problems that arise when the random variable $\underline{X}$ is grouped in two different ways are related to (a) same number of groups with same number of elements in each (b) same number of groups with different number of elements and (c) different number of groups. Since (a) and (b) give rise to same number of components in the corresponding first new GCVs, for testing purposes they can be tackled in a similar manner.

A random sample of size n, $\underline{X}_1,\ldots,\underline{X}_n$, is taken from $N_p(\underline{0},\Sigma)$. Let $\underline{Y}$ and $\underline{Z}$ be first new GCVs of order r and r-1 respectively. Under the approximations of Section 4.1 construct $\underline{Y}_i = (\underline{\beta}'_{Y1}\underline{X}^*_{1i},\ldots,\underline{\beta}'_{Yr}\underline{X}^*_{ri})'$ and $\underline{Z}_i = (\underline{\beta}'_{Z1}\underline{X}^{**}_{1i},\ldots,\underline{\beta}'_{Zr-1}\underline{X}^{**}_{r-1i})'$ for each $\underline{X}_i$, i=1,...,n where $\underline{X}' = (\underline{X}^{*'}_1,\ldots,\underline{X}^{*'}_r)$ and $\underline{X}' = (\underline{X}^{**'}_1,\ldots,\underline{X}^{**'}_{r-1})$ are the partitions of $\underline{X}$ into orders r and r-1 corresponding to $\underline{Y}$ and $\underline{Z}$ respectively. Then $(\underline{Y}_1,\ldots,\underline{Y}_n)$ and $(\underline{Z}_1,\ldots,\underline{Z}_n)$ can be considered, each, separately as a sample of GCVs corresponding to $\underline{Y}$ and $\underline{Z}$ respectively. But, $\underline{Y}_i$ and $\underline{Z}_i$ will still be dependent. This dependency can be avoided if a sub-sample $\underline{X}_1,\ldots,\underline{X}_m$ of size m is used to estimate Σ and then obtain $\underline{Y}_i$, i = 1,...,m, from this estimate and sub-sample and the remaining sample, $\underline{X}_j$, j = m+1,...,n is used to give an independent estimate of Σ and obtain $\underline{Z}_j$, j = 1,...,(n-m). A LRT can then be performed for the equality of the generalized variances of $\underline{Y}$ and $\underline{Z}$, using advantageously the independence of $\underline{Y}_i$ and $\underline{Z}_j$s. The same method can be used for more than two new GCVs. However, this procedure will be quite inefficient because Σ will be estimated on small sample sizes and so will

of course be the coefficient vectors. Further, the number of $\underline{Y}_i$s and $\underline{Z}_i$s will be greatly reduced. A modified method which uses the same number of GCVs $\underline{Y}_i$s and $\underline{Z}_i$s as that of the $\underline{X}_i$s, namely n, and where Σ is estimated using the entire sample of size n is proposed below. The method is also applicable in the case of several new GCVs.

Lemma 1: If $\underline{X}_1,\ldots,\underline{X}_n$ are i.i.d. and $\underline{Y}_1\ldots,\underline{Y}_n$ are i.i.d., $\underline{X}_i$ is independent of $\underline{Y}_j$, $i \neq j$, and $\underline{X}_i$ and $\underline{Y}_i$s have multivariate normal distributions, then $\underline{X} = \Sigma a_i \underline{X}_i$ and $\underline{Y} = \Sigma b_i \underline{Y}_i$ are independent if $\Sigma\, a_i b_i = 0$.

Proof. $\mathrm{Cov}(\underline{X},\underline{Y}) = \Sigma\Sigma\, a_i b_j \mathrm{Cov}(\underline{X}_i,\underline{Y}_j) = \Sigma\, a_i b_i \mathrm{Cov}(\underline{X}_i,\underline{Y}_i) = \mathrm{Cov}(\underline{X},\underline{Y}).0 = \underline{0}$

and hence the lemma follows from the normality of $\underline{X}$ and $\underline{Y}$.

Let $Y_j = \sum_{s=1}^{r} Y_{js}$, $Y^* = \sum_{1}^{n} a_j Y_j$; $Z_j = \sum_{s=1}^{r-1} Z_{js}$, $Z^* = \sum_{1}^{n} b_j Z_j$

Choose a_j, b_j such that $\Sigma\, a_j b_j = o$. Also if $\underline{Y} \sim N_r(\underline{0},\Sigma_{\rho_r})$, $Z \sim N_{r-1}(0,\Sigma_{\rho_{r-1}})$ then, $\mathrm{Var}(Y^*) = r\,(1+\overline{r-1}\rho_r)$ and $\mathrm{Var}\,(Z^*) = (r-1)\,(1+\overline{r-2}\rho_{r-1})$, where we have chosen a_j, b_j such that $\Sigma\, a_j^2 = \Sigma\, b_j^2 = 1$. Call the resulting variables Y^{**} and Z^{**}. Using the approximations of section 4.1 and by Lemma 1, in large samples Y^{**} and Z^{**} are independently distributed.

We want to test H_0: $\sigma_1 = \sigma_2$ where $\sigma_1 = |\Sigma_{\rho_r}|^{1/r}$ and $\sigma_2 = |\Sigma_{\rho_{r-1}}|^{1/(r-1)}$ against H_1:$\sigma_1 < \sigma_2$ or H_2:$\sigma_1 \neq \sigma_2$. For testing H_0 against H_2, we use LRT. Differentiating the Likelihood function with respect to ρ_r and ρ_{r-1} and equating the derivatives to zero, we have, $y^{**2} = r(1+\overline{r-1}\rho_r)$ and $z^{**2} = (r-1)(1+\overline{r-2}\rho_{r-1})$ (estimate of variances)

Then, under Ω, maximum of the Likelihood function $L(y^{**},z^{**})$ is $1/(y^{**}z^{**})$

Under H_0, we have to maximize the Likelihood function and then consider, $L^*(y^{**},z^{**}) = L(y^{**},z^{**}) + \lambda(\sigma_1 - \sigma_2)$, where λ is Lagrange's multiplier. The solution seems to need numerical or iterative methods. With these solutions, let Max L^{**} be $f(y^{**},z^{**})$, a function of y^{**} and z^{**}.

Reject H_0 iff $\eta = f(y^{**},z^{**})/(y^{**}z^{**}) < \eta_o$, a constant.

Under H_o, $-2 \ln \eta$ is approximately distributed as a χ^2 variable with 1 d.f.

So far, except the conditions (i) $\Sigma a_i^2 = \Sigma b_i^2 = 1$ and (ii) $\Sigma a_i b_i = o$, a_i, b_is are arbitrary. Since the statistic in LRT looks formidable, the power function of the above test may be difficult to compute. However, we may try to improve on the estimators, $\hat{\rho}_i$, i = r-1,r-2, obtained above when $L(y^{**},z^{**})$ was maximized under Ω, are unbiased estimates. We will try to find minimum variance unbiased estimators among the above class. This amounts to minimizing Σa_i^4 and in view of condition (i) above, subject to $\Sigma a_i^2 = 1$. The solution is easily found to be $a_i^2 = 1/n$,i = 1,...,n.

5.2 Tests for equality of multidimensional scatter of more than two new generalized canonical variables:

Here we will consider separately GCVs with same and different number of components, because some simplifications are available in the former case.

(a) GCVs with same number of components and with equi-correlations all > 0 or all < 0: Consider the case of s (say, even) GCVs with same number of components r and equi-correlation coefficients $\rho_i, .. i = 1,\ldots,s$ all >0 or all <0. Take s/2 independent sub-samples of same size and construct $(y_1^{**},y_2^{**}),\ldots,(y_{s-1}^{**},y_s^{**})$ as mentioned in section 5.1. Thus y_i^{**}s are independent normal variables. Then, the Likelihood function is given by L, where, writing $y_i \equiv y_i^{**}$,

$$\ln L = C - \tfrac{1}{2}\Sigma \ln\{r(1+\overline{r-1}\rho_i)\} + y_i^2/\{r(1+\overline{r-1}\rho_i)\}, \text{ where } C = (2\pi)^{s/2} .$$

Then, in the notations of Barlow et al (1972) letting,

$$X = \{1,\ldots,s\},\ g(i) = y_i^2,\ f(i) = \{r(1+\overline{r-1}\rho_i)\} \text{ and } w(i) = 1,\ i=1,\ldots,s$$

we have the

Lemma 1. The Isotonic regression g* maximizes L under H_1: The generalized variance of i^{th} GCV is greater than that of the j^{th} GCV, i>j.

Proof: Note first that since equi-correlations are all <0 (all >0) and ordering of the generalized variances of the new GCVs gives the same (reverse) ordering of the corresponding equi-correlation coefficients. With the above values of X, g(i), f(i) and w(i), we use Theorem 1.10 of Barlow et al quoted in our Appendix and the relevant notations. Let $\Phi(u) = -\ln u$. To complete the proof it suffices to note that maximizing ln L is equivalent to minimizing $\Sigma_i[\log f(i) + \{g(i)/f(i)\}]$ both subject to f isotonic, since the first and last terms of Δ_Φ do not involve f. (When $\rho_i > 0\ \forall i$, istonicity is w.r.t. reverse of the natural order on X and for computational purposes a formula is

$$g_i^* = \min_{s\le i}\ \max_{t\ge i}\ \sum_{j=s}^{t} y_j^2 .$$

When $\rho_i < 0\ \forall i$, min and max are interchanged in the above formula.)

Using the fact that under H_0 Max L is obtained by simply replacing $1+(r-1)\rho_i$ by y_i^2 in L and Lemma 1, the LRT statistic λ can be easily found.However, even under H_0 the distribution of the statistic seems to be complicated. In large sample, $-2\ln\lambda \sim \chi^{-2}$(see Barlow p. 198).

(b) GCVs with different number of components: In view of the above difficulties even in the special cases as in (a),we propose to use the approximation to the distribution of the generalized variance by Gamma distribution by Hoel (1937). Barlow et al. (p. 198) has considered a similar case and with a slight modification it follows that the LRT statistic , $-2\ln\lambda \sim \chi^{-2}$. We reject the hypothesis of equal generalized variances in favor of a given ordered relationship among them, if $-2\ln\lambda$ exceeds the upper α per-cent point of the distribution of χ^{-2}. Some tables are given in Barlow et al. For this test, however, see comments in Section 7.

6. DISCUSSIONS AND AN EXAMPLE

In this chapter new GCVs are found for the classical example due to Thurstone and Thurstone mentioned in Section 2.3 and studied by Horst, Kettenring, Seal and Gnanadesikan.The correlation matrix, R, for the 9 variables we consider here was given by,

$$R = \begin{pmatrix} & .636 & .126 & .059 & .626 & .195 & .059 \\ I & -.021 & .633 & .049 & .035 & .459 & .129 \\ & .016 & .157 & .521 & .048 & .238 & .426 \\ & & & & .709 & .050 & -.002 \\ & & I & & .039 & .532 & .190 \\ & & & & .067 & .258 & .299 \\ & & & & & & \\ & & & & & I & \end{pmatrix}$$

where I is the 3 x 3 identity matrix.

Here k =3 and p = 9. Since $R_d^{-1} = I$ (the 9 x 9 identity matrix) we need only to solve $|R - \hat{\lambda}^* I| = 0$, i.e. determine the characteristic roots, say $\hat{\lambda}_i^*$, of R and then find $\rho_i = (\hat{\lambda}_i^* - 1)/2$, i = 1,...,9. The coefficient vector for the s^{th} new GCV is the characteristic vector corresponding to $\hat{\lambda}^{*(s)}$, say, where the corresponding equi-correlation coefficient $\rho^{(s)}$ is the one which gives the s^{th} minimum value among $(1 + 2\rho_i)(1 - \rho_i)^2$, i = 1,...,9.; s = 1,...,9.

Table 1 exhibits the order of the first four new GCVs and the corresponding generalized variances. We note that (1) The new GCCs are computed quite easily e.g. $r^{(1)} = .745$. For the GENVAR method and several others discussed by Steel and Kettenring, however, both the GCCs and the coefficients of the GCVs for each stage need to be computed through extensive iterative methods. (2) For a negative value of r, i.e. $r = -.385$, we have the third new GCV and (3) Since $|\Sigma_r|$ for the third new GCV is as high as .45, it seems that the higher order new GCVs are redundant from practical considerations.

The computational aspect of the new GCVs is quite interesting. At the first stage, for the coefficients $\underline{a}_i^{(1)}$, one needs to solve a simultaneous system of non-linear equations

$$\underline{a}_i^{(1)\prime} S_{ii} \underline{a}_i^{(1)} = 1, \quad \underline{a}_i^{(1)\prime} S_{ij} \underline{a}_j^{(1)} = r^{(1)}, \quad i \neq j,\ i,j=1,\ldots,k$$

where $r^{(1)}$ has already been obtained (as above) from (2) of p.20. S can be first tranformed to a (block) correlation matrix R, as done by Horst for our above example. The conditions then become

$$\underline{b}_i^{(1)\prime} \underline{b}_i^{(1)} = 1, \quad \underline{b}_i^{(1)\prime} R_{ij} \underline{b}_j^{(1)} = r^{(1)}, \quad (\underline{b}_i^{(1)\prime} = \underline{a}_i^{(1)\prime} S_{ii}^{\frac{1}{2}}),\ i \neq j,\ i,j=1,\ldots,k.$$

Further, a polar transformation on $\underline{b}_i^{(1)}$ reduces by one the number of unknown variables in each set. The usual Gauss-Seidel method which determines one variable from each equation needs to be modified for our purpose. Suppose an initial value $\underset{o}{\underline{b}}^{(1)}$ of $\underline{b}^{(1)}$ is given. Usually a good initial value is $\underset{o}{\underline{b}}_i^{(1)} = \bar{\underline{b}}_i^{(1)} / \bar{\underline{b}}_i^{(1)\prime} \bar{\underline{b}}_i^{(1)}$, $i=1,\ldots,k$ where $\bar{\underline{b}}^{(1)\prime} = (\bar{\underline{b}}_1^{(1)\prime}, \ldots, \bar{\underline{b}}_k^{(1)\prime})$ is the eigen vector of R corresponding to $r^{(1)}$. We solve for $\underline{b}_2^{(1)}$ from $\underset{o}{\underline{b}}_1^{(1)\prime} R_{12} \underline{b}_2^{(1)} = r^{(1)}$ giving $\underset{1}{\underline{b}}_2^{(1)}$, for $\underline{b}_3^{(1)}$ from $\underset{1}{\underline{b}}_2^{(1)} R_{23} \underline{b}_3^{(1)} = r^{(1)}$ giving $\underset{1}{\underline{b}}_3^{(1)}$, and so on till finally for $b_1^{(1)}$ from $\underset{1}{\underline{b}}_k^{(1)} R_{k1} \underline{b}_1^{(1)} = r^{(1)}$ giving $\underset{1}{\underline{b}}_1^{(1)}$. $\underset{1}{\underline{b}}_i^{(1)}$ s are standardized to have unit

length, yielding the coefficient vector at the end of the first iteration as, using the same notation, $_1\underline{b}^{(1)}$ from $_o\underline{b}^{(1)}$. The procedure is then further iterated. Since $r^{(1)}$ is known previously, the iteration can be terminated at the n-th stage if

$$\sum_{i=1}^{k} \left| {}_n\underline{b}_i^{(1)\prime} R_{ii+1} \, {}_n\underline{b}_{i+1}^{(1)} - r^{(1)} \right| < \varepsilon$$

for some pre-assigned ε and the suffix k+1 is replaced by 1. It is reasonable to make a preliminary polar transformation on $\underline{b}_i^{(1)}$, each separately for the i-th set, i=1,...,k, as suggested earlier. The entire procedure outlined above can then be performed in terms of the transformed variables. The ZXMIN subroutine in IMSL can then be advantageously exploited to solve for $\underline{b}^{(1)}$. This procedure is applicable to coefficients for higher stage new GCVs also, $\underline{b}^{(2)}, \underline{b}^{(3)}, \ldots$ - only the equations defining the additional constraints at the corresponding stages have now to be also solved simultaneously.

For our example, a single iteration of the above procedure yielded

$$\underline{b}_1^{(1)\prime} = (.642691 \quad .619972 \quad .450091)$$

$$\underline{b}_2^{(1)\prime} = (.543281 \quad .733599 \quad .408263)$$

$$\underline{b}_3^{(1)\prime} = (.668885 \quad .663320 \quad .335559) \qquad \text{and}$$

$$\underline{b}_1^{(1)\prime} R_{12} \underline{b}_2^{(1)} = .745, \quad \underline{b}_2^{(1)\prime} R_{23} \underline{b}_3^{(1)} = .748, \quad \underline{b}_3^{(1)\prime} R_{31} \underline{b}_1^{(1)} = .749,$$

the value of $r^{(1)}$ being .745 as determined at the outset.

It is seen from this example that unlike several of the more reasonable methods studied in details by Kettenring, the new GCVs are quite straightforward and easy to compute. We have to merely find the characteristic roots and the standardized eigen vectors of a given matrix. From the

computational point of view the gain is significant. Also, we have already seen from Sections 4 and 5 several advantages of new GCVs from the point of view of statistical inference.

Table 1. $|\Sigma_r(s)|$ for ordered new GCVs

$\hat{\lambda}^*$	$r^{(s)}$	$\lvert\Sigma^{r(s)}\rvert$	new GCV
2.48985	.745	.16	first stage
2.16422	.582	.33	second stage
.23548	-.385	.45	third stage
1.61986	.310	.77	fourth stage

7. TOPICS FOR FUTURE RESEARCH

The tests based on the approximation to the distribution of the generalized variance by a gamma distribution may,with further research,be possibly improved upon since this approximation does not take into account the additional information of equi-correlation structure of the dispersion matrix. This leads to the estimation and exact distributional problems associated with the generalized variance corresponding to an equi-correlated dispersion matrix. Hence, new directions need to be sought for these problems and once the solutions are obtained, problems of statistical inference presented in this report can be considered with these modifications. A simulation study comparing the modified methods with the present one will be worth attempting. Properties of the proposed tests may also be considered in more details.

Relationships of GCVs with such important fields as Time Series Analysis, Regression Analysis, Prediction Theory, MANOVA, Discriminant Analysis, and Scaling and Factor Analysis are topics for future research. These are expected to achieve simplifications in terms of cost and analysis. The case with nominal variables (used extensively in Bio-Medical sciences) and a Ranking and Selection Procedure based on GCVs are interesting problems for further investigations.

Statistical inference associated with the previous GCVs and generalized variances in general, are important topics for future research.

For LRTs for generalized variances and their associated properties see Sen Gupta (1981c). Such tests are applicable to generalized canonical varibles obtained by minimizing the generalized variance, of general structure, as considered by Anderson (1958), Kettenring (1971), Steel (1951) and others.

Acknowledgements

It's a pleasure to acknowledge the helpful discussions I had with Professor C.R. Rao during this research. I am thankful to Professor T.W. Anderson and Professor R. C. Srivastava for their kind encouragements. Finally, I am grateful to Professor K. S. Lii for the computations required in the example.

APPENDIX

A.1 Theorem 1: Let A be positive definite and B be positive semi-definite. Then the roots of B in the metric of A are non-negative (where the roots of the equation $|\lambda A - B| = 0$ are called the characteristic roots of B in the metric of A). If B is also positive definite, then such roots are positive.

Proof: See Proposition 17, pages 581-582 of Dhrymes (1970).

A.2 Theorem 2: (Principle of Invariance of MLE) Let $\theta \in \Omega$ (an interval in a k-dimensional Euclidean space) and let $L(\theta)$ denote the Likelihood function - a mapping from Ω to the real line R. Assume that the MLE $\hat{\theta}$ of θ exists so that $\hat{\theta} \in \Omega$ and $L(\hat{\theta}) \geq L(\theta)$ for all $\theta \in \Omega$. Let f be an arbitrary transformation mapping Ω to Ω^* (an interval in an r-dimensional Euclidean space, $1 \leq r \leq k$). Then $f(\hat{\theta})$ is a maximum induced likelihood estimator of $f(\theta)$.

Proof: See Zehna (1966).

A.3 Theorem 3: (Theorem of Hurwitz) Let $\{f_n(x)\}$ be a sequence of analytic functions regular in a region G, and let this sequence be uniformly convergent in every closed subset of G. Suppose the analytic function $\lim_{n\to\infty} f_n(x) = f(x)$ does not vanish identically. Then if $x = a$ is a zero of $f(x)$ of order k, a neighbourhood $|x - a| < \delta$ of $x = a$ and a number N exist such that if $n > N$, $f_n(x)$ has exactly k zeros in $|x - a| < \delta$.

Proof: See page 22 of Szego (1939).

A.4 Definition (Isotonic regression): Let X be the finite set $\{x_1, \ldots, x_k\}$ with the simple order $x_1 \prec x_2 \prec \ldots \prec x_k$. A real valued function f on X is isotonic if $x, y \in X$ and $x \prec y$ imply $f(x) \leq f(y)$. Let g be a given function on

X and w a given positive function on X. An isotonic function g^* on X is an isotonic regression of g with weights w with respect to the simple ordering $x_1 \prec x_2 \prec \ldots \prec x_k$ if it minimizes in the class of isotonic function f on X the sum $\sum_{x \in X} [g(x) - f(x)]^2 w(x)$. When the weight function and the simple ordering are understood, we call g^* simply an isotonic regression of g.

Note: A binary relation "$\lesssim$" on X establishes a 'simple order' on X if

1. it is reflexive: $x \lesssim x$ for $x \in X$;
2. it is transitive: $x,y,z \in X$, $x \lesssim y$, $y \lesssim z$ imply $x \lesssim z$;
3. it is antisymmetric: $x,y \in X$, $x \lesssim y$, $y \lesssim x$ imply $x = y$;
4. every two elements are comparable: $x,y \in X$ implies either $x \lesssim y$ or $y \lesssim x$.

Algorithms for isotonic regression: Barlow et al (1972) have discussed in details an algorithm called the Pool-Adjacent-Violators algorithm for finding the isotonic regression g^*. They have also discussed a scheme in order to program this algorithm for a computer (see page 72 of Barlow et al). Using essentially this scheme Kruskal (1964) has written a program to carry it out as a part of a large program.

Theorem 4: (Theorem 1.10 of Barlow et al). For Φ convex, let

$$\Delta_\Phi(g(x),f(x)) \equiv \Delta(g,f) = \Phi(g) - \Phi(f) - (g - f)\phi(f)$$

where ϕ is the derivative of Φ at f, or if it does not have a derivative at f, $\phi(f)$ denotes any number between the left and the right derivative at f. If f is isotonic on X and if the range of f is in I then

$$\sum_x \Delta[g(x),f(x)]w(x) \geq \sum_{=x} \Delta[g(x),g^*(x)]w(x) + \sum_x \Delta[g^*(x),f(x)]w(x).$$

Consequently g^* minimizes $\sum_x \Delta[g(x),f(x)]w(x)$ in the class of isotonic f with range in I and maximizes $\sum_x \{\Phi[f(x)] + [g(x) - f(x)]\phi[f(x)]\}w(x)$.

The minimizing (maximizing) function is unique if Φ is strictly convex. A corollary which is useful for our purpose is

Corollary: Let $\psi_1,\ldots,\psi_p$ be arbitrary real valued functions and let $h_1,\ldots,h_m$ be isotonic functions on X. Then g^* minimizes $\sum_x \Delta[g(x),f(x)]w(x)$ in the class of isotonic functions f with range in I satisfying any or all of the side conditions

$$\sum_x [g(x) - f(x)]\psi_j[f(x)]w(x) = 0,\ j = 1,\ldots,p$$

$$\sum_x f(x)h_j(x)w(x) \geq \sum g(x)h_j(x)w(x),\ j = 1,\ldots,m\ .$$

Proof: See corollary in page 42 of Barlow et al.

BIBLIOGRAPHY

Aitken, M.A., Nelson, W.C. and Reinfurt, K.H. (1968): "Tests for correlation matrices," Biometrika 55, 327-334.

Anderson, T.W. (1958): An Introduction to Multivariate Statistical Analysis, John Wiley, New York.

Anderson, T.W. (1963): "Asymptotic theory for principal component analysis," Ann. Math. Stats. 34, 122-148.

Arvanitis, L.G. and Afonja, B. (1971): "Use of the generalized variance and the gradient projection method in multivariate stratified sampling," Biometrics 27, 119-127.

Barlow, R.E., Bartholomew, D.J., Bremner, J.M. and Brunk, H.D. (1972): Statistical Inference under order restrictions, John Wiley, New York.

Dhrymes, P.J. (1970): Econometrics: Statistical foundations and applications, Harper & Row Publishers, New York.

Eaton, M.L. (1967): "The generalized variance: testing and ranking problems," Ann. Math. Stat. 38, 941-943.

Frisch, R. (1929): Correlation and scatter in statistical variables," Nordic Stat. Jour. 8, 36-102.

Geisser, S. and Desu, M.M. (1962): "Predictive zero-mean uniform discrimination," Biometrika 55, 519-524.

Ghosh, S.P. (1958): "A note on stratified random sampling with multiple characters," Calcutta Stat. Assoc. Bull. 8, 81-90.

Giri, N.C. (1977): Multivariate Statistical Inference, Academic Press, New York.

Glesser, L.J. (1968): "On testing a set of correlation coefficients for equality," Biometrika 55, p. 513-517.

Gnanadesikan, R. and Wilk, M.B. (1969): "Data analytic methods in multivariate analysis," Multivariate Analysis II, ed. P.R. Krishnaiah, Academic Press, New York, 593-638.

Gnanadesikan, R., and Gupta, S.S. (1970): "A selection procedure for multivariate normal distribution in terms of generalized variances," Technometrics 12, 103-117.

Gnanadesikan, R. (1977): Methods for Statistical Data Analysis of Multivariate Observations, John Wiley, New York.

Goodman, M. (1966): "A measure of 'overall variability' in populations," Biometrics 22, 189-192.

Gupta, S.S. and Panchapakesan, S. (1969): "Some selection and ranking procedure for Multivariate Normal populations," Mult. An. II (ed. P.R. Krishnaiah), 475-505.

Han, C. (1968): "A note on the discrimination in the case of unequal covariance matrices," Biometrika 55, 586-587.

Hoel, P.G. (1937): "A significance test for component analysis," Ann. Math. Stats. 8, 149-158.

Horst, P. (1936): "Obtaining a composite measure from a number of different measures of the same attribute," Psychometrika 1, 53-60.

Hotelling, H. (1936): "Relations between two sets of variates," Biometrika 28, 321-377.

Huynh, H. and Feldt, L.S. (1970): "Conditions under which mean square ratios in repeated measurement design have exact F-distribution," Jour. of the Am. Stats, Assoc. 65, 1582-1589.

Kettenring, J.R. (1971): "Canonical Analysis of several sets of variables," Biometrika 58, 433-451.

Kruskal, J.B. (1964): "Nonmetric multidemensional scaling: a numerical method", Psychometrika 29, 115-129.

Kshirsagar, A.M. (1972): Multivariate Analysis, M. Dekker, New York.

Mathai, A.M. (1967): "Dispersion Theory," Estadistica 25, 271-284.

McKeon, J.J. (1965): "Canonical Analysis: some relations between canonical correlations, factor analysis, discriminant function analysis and scaling theory," Psychometric Monograph 13.

Milton, R.C. (1963): "Tables of the equally correlated Multivariate Normal probability integral," Technical Report No. 27, Univ. of Minnesota.

Press, S.J. (1972): Applied Multivariate Analysis, Holt, Rinehart & Winston, New York.

Rao, C.R. (1973): Linear Statistical Inference and its Applications, John Wiley & Sons, New York.

Seal, H.L. (1964): Multivariate Statistical Analysis for Biologists. Methuen, London.

Sen Gupta, A. (1980): "Generalized correlations in the singular case," Tech. Report 46, Contract N00014-75-C-0442, Dept. of Statistics, Stanford University.

Sen Gupta, A. (1981a): "Tests for standardized generalized variances of multivariate normal populations of possibly different dimensions," Tech. Report 50, Contract N00014-75-C-0442, Dept of Statistics, Stanford University.

Sen Gupta, A. (1981b): "On tests for the equicorrelation coefficient and the generalized variance of a symmetric standard multivariate normal distribution," To appear, Dept. of Statistics, Stanford University.

Sen Gupta, A. (1981c): "Generalized canonical variables." To appear, Encyclopedia of Statistical Sciences, vol. 3 (edited by N.L. Johnson, and S. Kotz) John Wiley, New York.

Steel, R.G.D. (1951). "Minimum generalized variance for a set of linear functions," Ann. Math. Stat. 22, 456-460.

Steyn, H.S. (1978): "On approximations for the central and noncentral distribution of the generalized variance," Jour. Am. Stats. Assoc. 73, 670-675.

Szego, G. (1975): "Orthogonal Polynomials," American Math. Society Colloquium publications, Vol. XXIII, Providence, R.I.

Thurstone, L.L. and Thurstone, T.G. (1941): "Factorial studies of invariance," Psychometric Monograph 2.

Timm, N.H. (1975): Multivariate Analysis with Applications in Education and Psychology, Brooks/Cole Publishing Company, California.

van der Vaart, H.R. (1965): "A note on Wilks' internal scatter," Ann. Math. Stat. 36, 1308-1311.

Vinograde, B. (1950): "Canonical positive definite matrices under internal linear transformation," Proceedings of the Am. Math. Soc., 1, 159-161.

Votaw, D.F. (1950): "Compound Symmetry Tests in the Multivariate Analysis of Medical Experiments," Biometrics 6, 259-281.

Wilks, S.S. (1967): "Multidimensional Statistical Scatter," Collected papers, Contributions to Math. Stats., ed. T.W. Anderson, 597-614.

Zacks, S. (1971): The Theory of Statistical Inference. John Wiley, New York.

Zehna, P.W. (1966). "Invariance of maximum likelihood estimation," Ann. Math. Stats. 37, 744.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE	READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER: 52 — 2. GOVT ACCESSION NO.: AD-A115 315	3. RECIPIENT'S CATALOG NUMBER:
4. TITLE (and Subtitle): On the Problems of Construction and Statistical Inference Associated with a Generalization of Canonical Variables	5. TYPE OF REPORT & PERIOD COVERED: Technical Report
	6. PERFORMING ORG. REPORT NUMBER:
7. AUTHOR(s): Ashis Sen Gupta	8. CONTRACT OR GRANT NUMBER(s): N00014-75-C-0442
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Department of Statistics, Sequoia Hall, Stanford University, Stanford, California	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: (NR-042-034)
11. CONTROLLING OFFICE NAME AND ADDRESS: Office of Naval Research, Statistics and Probability Program Code 436, Arlington, Virginia 22217	12. REPORT DATE: February 1982
	13. NUMBER OF PAGES: 49
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)	15. SECURITY CLASS. (of this report): UNCLASSIFIED
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE:

16. DISTRIBUTION STATEMENT (of this Report)

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

Also issued as Technical Report No. 168 - NATIONAL SCIENCE FOUNDATION GRANT MCS 78-07736 - Dept. of Statistics - Stanford University.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Multivariate analysis, canonical correlations, generalized variance, tests for independence.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Canonical correlation analysis provides a means by which the relation between two sets of variates can be assessed. The present study is concerned with an extension to K sets of variates. As in the case of two sets of variates, linear combinations of the variates are used, but now alternative criteria need be defined. The criteria developed are based on the generalized variance. Properties and tests using this criterion are obtained.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

TECHNICAL REPORTS

OFFICE OF NAVAL RESEARCH CONTRACT N00014-67-A-0112-0030 (NR-042-034)

1. "Confidence Limits for the Expected Value of an Arbitrary Bounded Random Variable with a Continuous Distribution Function," T. W. Anderson, October 1, 1969.

2. "Efficient Estimation of Regression Coefficients in Time Series," T. W. Anderson, October 1, 1970.

3. "Determining the Appropriate Sample Size for Confidence Limits for a Proportion," T. W. Anderson and H. Burstein, October 15, 1970.

4. "Some General Results on Time-Ordered Classification," D. V. Hinkley, July 30, 1971.

5. "Tests for Randomness of Directions against Equatorial and Bimodal Alternatives," T. W. Anderson and M. A. Stephens, August 30, 1971.

6. "Estimation of Covariance Matrices with Linear Structure and Moving Average Processes of Finite Order," T. W. Anderson, October 29, 1971.

7. "The Stationarity of an Estimated Autoregressive Process," T. W. Anderson, November 15, 1971.

8. "On the Inverse of Some Covariance Matrices of Toeplitz Type," Raul Pedro Mentz, July 12, 1972.

9. "An Asymptotic Expansion of the Distribution of "Studentized" Classification Statistics," T. W. Anderson, September 10, 1972.

10. "Asymptotic Evaluation of the Probabilities of Misclassification by Linear Discriminant Functions," T. W. Anderson, September 28, 1972.

11. "Population Mixing Models and Clustering Algorithms," Stanley L. Sclove, February 1, 1973.

12. "Asymptotic Properties and Computation of Maximum Likelihood Estimates in the Mixed Model of the Analysis of Variance," John James Miller, November 21, 1973.

13. "Maximum Likelihood Estimation in the Birth-and-Death Process," Niels Keiding, November 28, 1973.

14. "Random Orthogonal Set Functions and Stochastic Models for the Gravity Potential of the Earth," Steffen L. Lauritzen, December 27, 1973.

15. "Maximum Likelihood Estimation of Parameters of an Autoregressive Process with Moving Average Residuals and Other Covariance Matrices with Linear Structure," T. W. Anderson, December, 1973.

16. "Note on a Case-Study in Box-Jenkins Seasonal Forecasting of Time series," Steffen L. Lauritzen, April, 1974.

TECHNICAL REPORTS (continued)

17. "General Exponential Models for Discrete Observations," Steffen L. Lauritzen, May, 1974.

18. "On the Interrelationships among Sufficiency, Total Sufficiency and Some Related Concepts," Steffen L. Lauritzen, June, 1974.

19. "Statistical Inference for Multiply Truncated Power Series Distributions," T. Cacoullos, September 30, 1974.

Office of Naval Research Contract N00014-75-C-0442 (NR-042-034)

20. "Estimation by Maximum Likelihood in Autoregressive Moving Average Models in the Time and Frequency Domains," T. W. Anderson, June 1975.

21. "Asymptotic Properties of Some Estimators in Moving Average Models," Raul Pedro Mentz, September 8, 1975.

22. "On a Spectral Estimate Obtained by an Autoregressive Model Fitting," Mituaki Huzii, February 1976.

23. "Estimating Means when Some Observations are Classified by Linear Discriminant Function," Chien-Pai Han, April 1976.

24. "Panels and Time Series Analysis: Markov Chains and Autoregressive Processes," T. W. Anderson, July 1976.

25. "Repeated Measurements on Autoregressive Processes," T. W. Anderson, September 1976.

26. "The Recurrence Classification of Risk and Storage Processes," J. Michael Harrison and Sidney I. Resnick, September 1976.

27. "The Generalized Variance of a Stationary Autoregressive Process," T. W. Anderson and Raul P.Mentz, October 1976.

28. "Estimation of the Parameters of Finite Location and Scale Mixtures," Javad Behboodian, October 1976.

29. "Identification of Parameters by the Distribution of a Maximum Random Variable," T. W. Anderson and S.G. Ghurye, November 1976.

30. "Discrimination Between Stationary Guassian Processes, Large Sample Results," Will Gersch, January 1977.

31. "Principal Components in the Nonnormal Case: The Test for Sphericity," Christine M. Waternaux, October 1977.

32. "Nonnegative Definiteness of the Estimated Dispersion Matrix in a Multivariate Linear Model," F. Pukelsheim and George P.H. Styan, May 1978.

TECHNICAL REPORTS (continued)

33. "Canonical Correlations with Respect to a Complex Structure," Steen A. Andersson, July 1978.

34. "An Extremal Problem for Positive Definite Matrices," T.W. Anderson and I. Olkin, July 1978.

35. " Maximum likelihood Estimation for Vector Autoregressive Moving Average Models," T. W. Anderson, July 1978.

36. "Maximum likelihood Estimation of the Covariances of the Vector Moving Average Models in the Time and Frequency Domains," F. Ahrabi, August 1978.

37. "Efficient Estimation of a Model with an Autoregressive Signal with White Noise," Y. Hosoya, March 1979.

38. "Maximum Likelihood Estimation of the Parameters of a Multivariate Normal Distribution,"T.W. Anderson and I. Olkin, July 1979.

39. "Maximum Likelihood Estimation of the Autoregressive Coefficients and Moving Average Covariances of Vector Autoregressive Moving Average Models," Fereydoon Ahrabi, August 1979.

40. "Smoothness Priors and the Distributed Lag Estimator," Hirotugu Akaike, August, 1979.

41. "Approximating Conditional Moments of the Multivariate Normal Distribution," Joseph G. Deken, December 1979.

42. "Methods and Applications of Time Series Analysis - Part I: Regression, Trends, Smoothing, and Differencing," T.W. Anderson and N.D. Singpurwalla, July 1980.

43. "Cochran's Theorem, Rank Additivity, and Tripotent Matrices." T.W. Anderson and George P.H. Styan, August, 1980.

44. "On Generalizations of Cochran's Theorem and Projection Matrices," Akimichi Takemura, August, 1980.

45. "Existence of Maximum Likelihood Estimators in Autoregressive and Moving Average Models," T.W. Anderson and Raúl P. Mentz, Oct. 1980.

46. "Generalized Correlations in the Singular Case," Ashis Sen Gupta, November 1980.

47. "Updating a Discriminant Function on the Basis of Unclassified Data," G.J. McLachlan and S. Ganesalingam, November 1980.

48. "A New Proof of Admissibility of Tests in the Multivariate Analysis of Variance." T.W. Anderson and Akimichi Takemura, January, 1981.

49 "A-Optimality for Regression Designs," N. N. Chan, January 1981.

50. "Tests for Standardized Generalized Variances of Multivariate Normal Populations of Possibly Different Dimensions," Ashis Sen Gupta, February 1981.

51. "A Linear Theory for Noncausality," J.P. Florens and M. Mouchart, November 1981.

52. "On the Problems of Construction and Statistical Inference Associated With a Generalization of Canonical Variables," Ashis Sen Gupta, February 1982.


FILMED
7-8